TIGHT BINDING BOOK

A SURVEY OF PHYSICS

BY

FREDERICK A. SAUNDERS
PROFESSOR OF PHYSICS, HARVARD UNIVERSITY

LONDON
G. BELL & SONS, LTD
NEW YORK: HENRY HOLT AND COMPANY

PREFACE

Our modern life is so profoundly affected by the products of scientific thought and invention that no intelligent person can afford to remain wholly ignorant of the results and methods of this sort of study. Among the sciences, physics is remarkable for the fundamental nature of the phenomena with which it deals. Physical investigation leads to the formulation of "laws" which are condensed statements expressing very general relationships governing the behavior of inanimate matter, but affecting living beings also. The reader must not think that because these laws are so fundamental they are also remote from the details of the real world. A very little acquaintance with them discloses their essentially useful and practical nature, and one soon begins to appreciate them for their compactness and their wide applicability.

The reader who makes a survey of the field covered by this science will be introduced to some of the mysteries of nature, as well as to many ingenious inventions of mankind. He should consider first the ideas of mechanics, since so much of the behavior of matter cannot be understood until these general principles have become familiar. As soon as he begins to examine matter in its common states he should meet with the newer, and presumably better, conceptions usually referred to as "modern" physics. These will prove useful and stimulating, perhaps at times even irritating. He should, before he completes the survey, be given a few telescopic glimpses of the unexplored territory into which research is continually attempting to penetrate. In order to make room for such ideas, some of the more or less obsolete parts of the subject are omitted in this book, or condensed. The material selected has been chosen on the basis of the importance and interest of the ideas, though it is obvious that no two persons will agree exactly as to how this should be done.

The text has been written primarily for the college student, who, on account of his maturity, will be content with a quite compact treatment of the more elementary portions of the subject. It is

expected that the illustrations and examples here given will be liberally supplemented by class and laboratory experiments. Nothing can take the place of first-hand contact with physical phenomena; the more chances the student has to do things with his own hands the better. With occasional omissions, guided by the judgment of the instructor and the indications in the text itself (small print, etc.), most of the material here presented can be included in an ordinary college course. The better prepared or more gifted students are, of course, capable of going farther than the others, and should not be deprived of the opportunity.

No one can master a subject like physics by the mere act of reading it. The doing of problems constitutes an important part of the learning process. An attempt has been made here to encourage a thoughtful consideration of the subject by a choice of problems dealing with common things, but often in such a way that they cannot be solved by mere substitution in a formula. Several subjects are thus introduced which make good material for class discussions. No great store of mathematical knowledge is required. It is assumed that the reader has, or can quickly acquire, a knowledge of the metric system of weights and measures, and that he is acquainted with the mysteries of elementary algebra, geometry, and arithmetic. For many the latter constitutes the chief "difficulty" with physics. A page in the appendix supplies all the trigonometry that is needed. Beyond this the reader's equipment is supposed to include an inquiring mind, ordinary human powers of observation, and an average acquaintance with modern civilization.

The writer is probably indebted to nearly all the text-books which he has ever read or used in his teaching. Other people's ideas are stored away in the back of one's mind and spring forth later with every aspect of originality. He wishes to thank all those to whom he is thus unknowingly under obligations. He is, however, aware of much indebtedness to many friends who have contributed directly to the preparation of this work. To his colleagues, Professors N. H. Black and F. H. Crawford, who have read the manuscript and made many valuable criticisms, he owes his special thanks. Whatever clarity of argument or smoothness of diction the book possesses he owes largely to his wife, whose thoughtful care in going over the whole work has saved him from many an error. To his colleagues, Professors E. C. Kemble, E. L. Chaffee,

and H. E. Clifford, and to Professor N. A. Kent of Boston University, he is indebted for their kindness in revising certain chapters, on which their judgment has been invaluable. Many of his other colleagues, notably Professors T. Lyman, P. W. Bridgman, G. W. Pierce, H. H. Plaskett, C. G. Dawes, and R. deC. Ward have contributed ideas, criticisms, or material for illustrations, as have also some of the graduate students in the Jefferson Physical Laboratory, especially Dr. C. D. Reid and Messrs. A. F. Birch and A. W. Parkes. Other scientific men and institutions have kindly contributed illustrative material: the Dominion Observatory at Ottawa, Dr. C. G. Abbot, Secretary of the Smithsonian Institution of Washington, Dr. J. A. Fleming of the Department of Terrestrial Magnetism of the Carnegie Institution, Dr. R. W. G. Wyckoff of the Rockefeller Institute for Medical Research, and Professors D. C. Miller of the Case School of Applied Science, S. R. Williams of Amherst College, W. D. Harkins of the University of Chicago, C. L. Norton of the Massachusetts Institute of Technology, and A. L. Foley of Indiana University.

Many industrial organizations have kindly sent photographs or other material. Special thanks are due to many branches of the General Electric Company, to the Bell Telephone Laboratories, the Westinghouse Electric and Manufacturing Company, the S. Morgan Smith Company, the Thwing Instrument Company, the Cutler-Hammer Company, the Weston Electrical Instrument Company, the Central Scientific Company, the Bausch and Lomb Optical Company, the Allis-Chalmers Manufacturing Company, the Eastman Kodak Company, and the Electric Storage Battery Company.

The courses in Harvard College which cover the ground outlined in this book include laboratory work the directions for which are given in a Physical Laboratory Manual, by F. A. Saunders and F. H. Crawford (Harvard University Press).

F. A. Saunders

Jefferson Physical Laboratory,
Harvard University.

CONTENTS

MECHANICS

HEAT

SOUND

ELECTRICITY AND MAGNETISM

LIGHT

APPENDIX

MECHANICS

CHAPTER 1

LIQUIDS AT REST

The subject of physics has to do with the study of nearly all inanimate nature, and underlies all the other sciences, as well as the scientific developments in engineering, agriculture, medicine, the fine arts, etc. Those who know physics well are best able to say beforehand what practical plan will be most likely to work out successfully in a given situation, even though nothing just like it may ever have been tried before. Knowledge of this sort comes from a thorough understanding of the principles of science and can be acquired only by a consideration of a wide range of subjects, and the intelligent accumulation of facts, laws and methods of attack. There is very little chance about such knowledge. It is only by sound, careful thinking, continually put to the test by the solution of problems, that anyone can attain to a mastery of such a subject.

It is hard to classify the vast body of knowledge included under the head of physics. Not only are the phenomena of inanimate nature almost infinitely various, but they have a puzzling way of being interrelated, so that some of those which once appeared to be quite independent are now known to be due to identical causes. All the sciences, and the parts of any one science overlap. Still, for convenience the conventional divisions of physics into mechanics, heat, sound, magnetism, electricity, and light may be retained in such a book as this. The fact must be remembered, however, that there are no boundaries between these divisions. They no more interrupt the continuity of the subject than milestones break up a road; but the traveler finds them a convenience, and perhaps derives a certain satisfaction from noting them as they go by.

Since we cannot begin, as we should like to, simultaneously in all parts of physics, but must choose some one subject as a start-

ing point, we turn first to a consideration of some of the phenomena furnished by liquids.

So many interesting, practical and familiar matters have to do with water and our experiences on or in it that this makes a very good topic through which to approach the main body of our subject. As these first considerations are straightforward and comparatively simple, it is not hard to formulate general statements about the behavior of liquids, and these can in turn be applied in many useful ways.

Pressure and density. In this work we must make use of a number of unfamiliar words, and use a few familiar ones in much more definite and exactly fixed senses than usual. At the very first we must talk about forces, weights, pressures, etc. Common experience teaches us something about the nature of force and weight; for instance, we know that they are similar, and that both are commonly measured in terms of tons, pounds, kilograms, grams, etc.,[1] depending on the system of units being used. A discussion of these terms will be found in Chapter 5, and for the present we can get along without more formal introduction to them. The term ***pressure*** (p) we must confine rigidly to one definite meaning, namely the amount of force (F) per unit area; or, in algebraic symbols, if A is any area, and F the whole force acting on it, $p = F/A$. Pressures are usually expressed in pounds per square inch (lbs./in.^{2}), or in grams per square centimeter (gr./cm.^{2}). The ***density*** of a body is the mass (see p. 61) of unit volume of the substance, or the mass of a body divided by its volume; $d = m/v$. The ***specific gravity*** is the ratio of the density of the body to that of water. In the metric system the density and specific gravity are expressed by the same number, since the density of water is taken as 1. This same number expresses the specific gravity in the English system; but, in the latter system the density is 62.4 times the specific gravity, since a cubic foot of water weighs 62.4 pounds.

Facts about liquid pressures. Several general conclusions about liquid pressures may be reached without trouble. To begin with, let us imagine a quantity of liquid at rest in a vessel. Any particle of it must be acted on by equal and opposite forces in all

[1] The reader is assumed to be reasonably familiar with the metric system of weights and measures. A table is given in the appendix of the comparative values of the common units.

directions; for, if not, it would move in the direction of the greatest force. Thus ***the force on the particle and hence also the pressure must be the same in all directions at a point.***

Also, the force exerted by a liquid against the wall of a vessel must be directed ***perpendicularly to the wall;*** for, if it were at any other angle there would be nothing to prevent the liquid from flowing along the wall in the direction determined by the force. Water will, as we all know, squirt out of a hole (Fig. 1–1) in the wall of a vessel no matter in what direction the stream has to start, and the first part of this jet will always be at right angles to the wall in which the hole exists. A statement similar to that above holds for a ***free surface;*** the surface sets itself perpendicular to the force acting, which is commonly that of gravity; hence the surface is usually horizontal. Sometimes, however, as happens near the edge of a glass vessel partly filled with water, we can see that the surface is curved; but here there are at least two forces, weight and surface tension (see p. 128), acting at once, and as before the surface at every point sets itself at right angles to the combined effect of the forces.

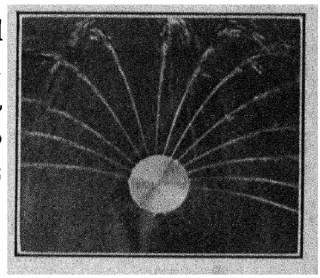


FIG. 1–1
Little streams of water forced out of holes in a round brass vessel, starting always perpendicularly.

Pressure and depth. It is easy to see also that ***the pressure must be proportional to the depth*** below the surface. Let us consider a cylindrical portion of the water in a vessel, as in Fig. 1–2, in contact with the surface, the cylinder having imaginary walls, or real ones of extreme thinness. The liquid inside this cylinder is at rest, like every other part of the liquid. Any force acting against its sides must be balanced by an equal and opposite force. Such forces must be at right angles to the cylinder walls, and therefore horizontal. No horizontal force is capable of lifting anything, or of exerting any other direct vertical effect. Therefore, since the weight of the cylinder ***is*** supported, these forces against the walls have nothing to do with it, and there must be a vertical force [illegible] the bottom of the cylinder just equal to

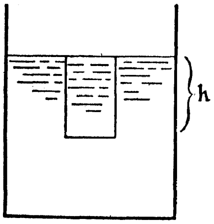


FIG. 1–2

the weight itself. The volume of the cylinder is its area A times its height h. If its density is d, its weight [1] will be Ahd. Since p is the force per unit area, the upward force on the bottom of the cylinder is pA. Hence the product pA is numerically equal to Ahd, or the pressure itself is measured by the number hd. This relation will be true for all depths, unless the density changes with the depth, which does not occur to any notable extent with ordinary liquids. The equation implies that the pressure is the same at the same depth in any shape or size of vessel, or in the ocean itself; that the depth is to be measured ***vertically*** even in cases (as for instance water in a slanting pipe) where this distance will not all lie within the liquid itself. Also, it refers only to pressures that are due to the weight of the liquid itself. Additional pressure put on the liquid from above must be considered separately (p. 8).

Force on the sides of vessels. To find the total force exerted on the side of a vessel full of liquid, we must allow for the fact that the pressure varies with the depth. Since it varies uniformly, the pressure halfway down is an average value, and this we may imagine to exist uniformly all over the wall in question. The total force is then this average pressure multiplied by the area of the wall.

Examples. Now let us see how these general remarks lead to practical results.

To find the pressure 1 *ft. under water* in pounds per square inch. A cubic foot of water weighs 62.4 lbs. It rests on its base of 144 sq. in. There rests therefore *on each square inch* a weight of 62.4/144, or 0.433 *lbs. at the depth of* 1 *ft.* From this it follows that a depth of 1/0.433, or 2.31 *ft.*, *corresponds to a pressure of* 1 *lb./in.*^{2}

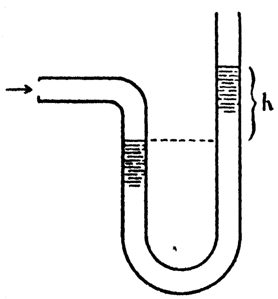


FIG. 1–3
Pressure gauge or manometer

To find the pressure in grams per square centimeter 1 *meter down.* A cubic centimeter of water weighs (very approximately) 1 gr., and exerts on its base a pressure of 1 gr./cm.^{2} A column 100 cm. deep would then exert a pressure of 100 gr./cm.^{2}

To find the pressure indicated by a gauge. Pressure gauges, sometimes called manometers, are often made of a bent glass tube, filled with a liquid as shown in Fig. 1–3. If the pressure is [illegible]

[1] ... here taken as num[illegible]

occurs for instance in the case of city gas pressures, water is often used as the liquid. If the pressures are high, mercury is better, as the liquid column is then 13.6 times shorter, and thus more convenient. This arises from the fact that mercury is approximately 13.6 times heavier than water. In such a gauge the pressure of the air (see Chapter 2) acts equally on both sides and balances out. A pressure measured by such devices may be called briefly "a pressure of 7 cm. of water" or "of 76 cm. of mercury," etc.; meaning, of course, a pressure equal to that found at the bottom of a column of water 7 cm. deep, or of mercury 76 cm. deep. This sort of scientific slang is regarded as allowable, since it saves time.

As an example, let us find the pressure in grams per square centimeter which is capable of supporting one side of a mercury manometer at a height of 76 cm. above the other. This pressure is equal to hd, which is 76×13.6 or 1033.6 gr./cm.^{2}

PROBLEMS

1. A water tank is made of wood, strengthened with iron hoops to hold it together. How should these hoops be spaced to relieve the strain evenly?

2. The liquid in a coffee pot weighs far more than the small amount in the spout. Why is the coffee in the spout not forced out by this great weight?

3. A faucet has an opening of $\frac{1}{3}$ sq. in. area. The water pressure is 100 lbs./in.^{2} How much force must you exert with your thumb against the opening to hold back the flow of water?

4. A dormitory is to be built on ground where the water pressure is 60 lbs./in.^{2} How much pressure will there be on the fourth floor of this building, assuming it to be 48 ft. above the ground?

5. A hole of 5 cm.^{2} is discovered in the bottom of a ship, 3 m. below the surface of the water. How much force will be required to resist the flow of water into the ship? (Solve for cases both of fresh water and of sea water of density 1.03.)

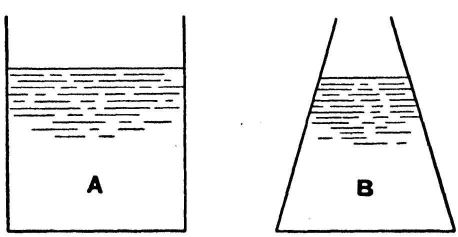


FIG. 1–4

6. Find the force exerted on the bottom of a vessel A (Fig. 1–4) by water which fills it to a depth of 10 cm. if the area of the base is 100 cm.^{2}

7. Find the force on the bottom of vessel B, filled to the same depth as A, and having the same area of base.

8. In the case of vessel B, how does the liquid manage to press on the bottom of the vessel with more than its own weight? (Consider a man in a low attic room with sloping roof, and think how he could exert more force on the floor than his own weight.)

9. Find the whole force on the side of an aquarium, 20 cm. high and 30 cm. wide, if it is filled to the top with salt water, whose density is 1.03 grams per cubic centimeter.

10. A gate on a small canal is 20 ft. wide and 16 ft. deep under water. Find the total force on one side of this gate. If the gate connects directly with a very large lake at the same level, compare the forces on the two sides of the gate.

11. A city has buildings 200 ft. high. What water pressure must be maintained at the ground level in order to protect these buildings properly from fire?

12. A uniform U-tube, open at both ends, is partly filled with water. Into one arm some oil (specific gravity 0.8) is gently poured, so that it stands as a column 10 cm. high on top of the water in that arm. How high will the water surface then be on the other side, measured from the top of the oil surface; and how high would it have been if the tubing on the water side had been half as wide?

13. A rectangular water storage tank is 3 ft. wide, 2 ft. high and 10 ft. long. A pipe 1 in. in diameter and 9 ft. long is screwed into a hole in the top of the tank, so that it stands vertical. Water is run into the tank until both tank and pipe are full. Find the pressure on the floor of the tank and the total upward-force on the top of the tank.

External pressures on liquids. What has already been said about pressures in liquids shows that if more water is poured into a vessel the level rises to a new height and the extra weight produces an ***equal increase in pressure*** everywhere throughout the water, this being added at every point to the pressure that existed there before. If extra pressure is applied by other means, by air pressure for instance, or by pressure from a piston, the result is the same. Pascal [1] stated this as a general principle: ***extra pressure applied to liquids from outside is felt undiminished in all parts of the liquid.*** Applications of this principle are common; one will be given.

[1] Blaise Pascal (1623–1662), religious philosopher and mathematician; one of the founders of the dynamics of liquids. He was chiefly interested in religious ideas, but was original as a mathematician, and carried out a few practical experiments, especially on the pressure of the air.

The hydraulic press. The hydraulic press offers an interesting example of a device for increasing force, which we shall soon see is a common property of several types of machines. In Fig. 1–5, which shows the principle of the press, a moderate force F is applied to a small piston of area a, which can move downward in a tall cylinder. It drives the liquid below it, water or oil, over into a large cylinder of area A and lifts the large piston of this cylinder, on which there rests a heavy weight W. In a real hydraulic press the weight W is replaced by the object to be compressed, which is placed between the large piston and an unyielding framework above. The pressure exerted by the force F is F/a, according to the definition of pressure. This is transmitted to all parts of the liquid in the press, and is usually so large that the pressure produced by the weight of the liquid itself can be neglected in comparison. This means that the bottom of the large piston has a pressure of F/a on each unit of its area, and thus it is pushed upward by a total force of FA/a. If this supports the weight W, then $W/F = A/a$, or W is larger than F in the same proportion that A is larger than a. It is not difficult to arrange to have the large piston ten times as large as the small one in diameter, or 100 times in area. Thus we increase the force 100 times with such a press. Since as much water must flow out of the narrow cylinder as flows into the wide one, the narrow piston in this example must move a hundred times further than the wide one; or, in other words, the motion of the large piston is usually small. There are other machines which multiply force, such as levers, for instance. To all of these a similar statement applies; if a small force moves the point at which it acts through a large distance, the large force moves through a proportionately small distance. The product of the force and the distance is the work done (p. 79) and this remains constant.

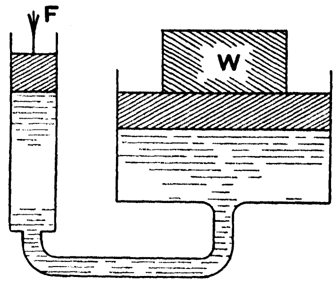


FIG. 1–5
The hydraulic press

Bridgman[1] has applied Pascal's principle to the production of pressures of over 300,000 pounds per square inch and has examined at these huge pressures many of the unusual properties

[1] P. W. Bridgman, Professor of Physics, Harvard University.

which matter then possesses. Thus water can be compressed into three-fourths of its usual volume and made to freeze when it is very hot, copper flows like soft butter when a high pressure is urging it and hydrogen goes freely through a two-inch thickness of steel.

The principle of Archimedes. Bodies when immersed in a liquid displace it and are then pushed upward by it with a force equal to the weight of the displaced liquid. This is called the principle of Archimedes [1] and was probably stated by him about 250 B.C. It is easy to see why it must be true. The cylinder in Fig. 1–2 was made of the liquid itself. If it had been made of some solid body the upward force on it would have been the same while it was in the same position. This upward force is the weight of the liquid in the cylinder, or the weight of the liquid displaced by it. If the body had been a heavy one, this force would not have sufficed to keep it afloat, but it would still have suffered an apparent loss of weight equal to the upward force exerted by the water on the body. If the body were lower down in the water, the circumstances would be unchanged, for the pressure it would then experience on top tending to sink it would be balanced by an equal extra pressure below tending to float it; and it would appear to lose weight by the same amount as before.

The principle of Archimedes can be applied to the otherwise awkward problem of finding the specific gravity of an irregular solid. If it sinks in water, its loss of weight under water can be found by hanging it from a balance arm into a vessel of water. This loss is equal to the weight of the displaced water. In the metric system this is numerically equal to its volume in cubic centimeters, and from this the specific gravity readily follows if we know the weight of the body in air.

A body light enough to float will rise from the position of the cylinder in Fig. 1–2 until it displaces only that amount of liquid whose weight is equal to its own. Since weight is proportional to density multiplied by volume, this product must be the same

[1] Archimedes (287–212 B.C.); Greek mathematician and inventor. In common with others of his time he regarded experimental work as of little importance, but traditions of his inventions survived as popular tales and he was credited with doing most unlikely things, such as setting Roman ships on fire by concentrating the sun's rays on them with mirrors. He himself regarded his geometrical theorems as his greatest achievements.

for the floating body and for the liquid which it displaces; from which it follows that

$$\text{density of body} \times \text{volume} = \text{density of liquid} \times \text{its volume}$$

or

$$\frac{\text{density of body}}{\text{density of liquid}} = \frac{\text{volume displaced}}{\text{volume of body}}.$$

This relation is of use in calculating the proportion of any floating body which is above or below the surface.

A ship floats because it displaces a large amount of water. The principle of Archimedes enables us to calculate the amount of the water displaced when a heavy cargo is put into the ship, and the horizontal area the ship must have at the water line in order not to sink too far down into the water on account of the added weight.

Example. A rectangular coal barge is 15 × 60 ft. and has vertical sides. If it can sink with safety 4 ft. deeper into the water when full than when empty, how many tons of coal can it carry? Maximum volume of water displaced is 15 × 60 × 4, or 3600 cu. ft. This weighs 3600 × 62.4 lbs., or 224,640 lbs. which is 112.3 tons.

TABLE I

Density of Solids and Liquids

In the metric system, always given in grams per cubic centimeter. To obtain pounds per cubic foot, multiply by 62.4

Metals		*Other Solids*	
Aluminum	2.7–2.8	Wood, white pine	0.35–0.50
Copper	8.9	maple	0.62–0.75
Gold	19.3	oak	0.60–0.90
Iron	7.85–7.90	Common rocks	2.5–3.0
Lead	11.3	Common glass	2.4–2.8
Magnesium	1.74	Ice	0.917
Platinum	21.5		
Silver	10.5	*Liquids*	
Tin	7.3	Sea water	1.03
Zinc	7.1	Machine oil	0.88–0.92
		Gasoline	0.74
Alloys		Alcohol (ethyl)	0.807
Brass (yellow)	8.4	Alcohol (denatured)	0.83
Duralumin	2.79	Milk	1.028–1.035
Magnalium	2.0	Mercury	13.6
Steel	7.6–7.9		

PROBLEMS

1. Find the volume both in cubic inches and in cubic centimeters of 1 lb. weight (454 gr.) of each of the following materials, gold, iron, pine wood. Use Table I.

2. Find the density and specific gravity of each of the following objects in both English and metric systems: (*a*) a feather pillow whose weight is 1 lb. and volume $\frac{3}{4}$ cu. ft.; (*b*) a log whose weight is 200 lbs. and volume 5 cu. ft.; (*c*) a block of metal whose weight is 150 lbs. and volume $\frac{1}{3}$ cu. ft. One cubic foot may be taken as 28,300 cubic centimeters.

3. How large is a square sheet of zinc whose thickness is 1 mm. and weight 900 gr. (approximately 2 lbs.). How big a square would it make if rolled out to one-fourth the thickness?

4. A stone weighs 100 kilograms (kg.) in air, and has a specific gravity of 2.5. How much force does it take to lift it under water?

5. A stone thrown into the sea sinks. Will it descend to the bottom, however deep? Under what conditions might it go only part way?

6. A fish just caught is hung on a spring balance and weighs 3 lbs. What does the fish appear to weigh if hung so that it is under water but the balance is above water? Why?

7. What is the volume of a 160-lb. man in cubic feet if he can just float with the tip of his nose out of water?

8. A uniform stick of wood, 100 cm. long and of density 0.6 gm./cm.,^{3} is made to float vertically in water. What length is submerged? How would it float in a liquid of density 0.9? If a scale were marked on it, the reading of the position of the surface on the scale could be made to give the density directly. Such an instrument is a simple form of *hydrometer* (Fig. 1–6).

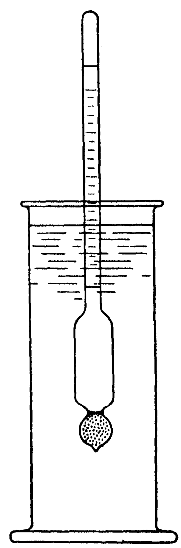


FIG. 1–6
A hydrometer

9. The piston under a barber's chair is 4 in. in diameter. If the weight of the chair and its occupant is 250 lbs., how much pressure (neglecting friction) is required to raise the chair. If this pressure is exerted by means of a plunger half an inch in diameter, how much force must be exerted?

10. A canoe with two people in it has an area of cross-section at the water level of 12 sq. ft. If we assume that the sides are vertical above the surface of the water, how much weight can be put into the canoe without sinking it more than 4 inches lower in the water?

11. An iceberg (density 0.9) has a total volume of 10,000 cubic meters. If we assume that it is floating in water of density 1.02, how many cubic meters of the ice are above water?

12. The specific gravity of brass is 8.4. A brass casting weighs 168 gr. in air and 143 gr. in water and has a cavity inside. How large is the cavity?

13. A ship is to carry 2000 tons of cargo without sinking more than 5 ft. deeper into the water than when empty. It is to sail in salt water of specific gravity 1.03. What size of ship is required, measured in square feet of area at the water line?

14. A raft is made of 6 logs, each 12 ft. long and 1 ft. diameter. The wood has a specific gravity of 0.6. What is the greatest weight in lbs. that this raft will support?

15. How much cork, of specific gravity 0.25, must be used to make a life preserver to support one-fifth of a man's body out of water? Assume the man's weight to be 60 kg. and his average density to be such that he can just barely float in water.

16. What proportion of an iron ball (specific gravity 7.8) will be submerged when it is floating on mercury (specific gravity 13.6)?

17. A piece of glass, of density 2.5, is hung by a thread from a balance and found to weigh 50 gr. in air. It appears to weigh only 28 gr. when it is allowed to hang freely immersed in a salt solution. Find the density of the solution.

18. Fishermen use lead weights to sink the bottom of their nets, with cork above to keep them upright under water. Find how a vertical string under water is pulled when 10 lbs. of lead (specific gravity 11.0) is fastened to the bottom of it, and 2 lbs. of cork (specific gravity 0.25) to the top of it, the cork also being completely under water. Will the string go up or down? What will be the final tension (force) in the string when the motion is over?

19. Archimedes developed his celebrated principle when he was asked (so the story goes) to find out whether the king's crown was of pure gold, or of gold mixed with silver. If it had been nine-tenths gold by weight, and one-tenth silver, what should he have found for its specific gravity, assuming no shrinkage of the materials when melted together? (Sp. gr. gold, 19.3; silver, 10.5.)

20. A piece of material of density 3 is being weighed accurately on a sensitive balance. Brass weights are used to counterbalance it, of density 8.4. The material appears to weigh exactly 300.00 gr. What buoyant effect would the air have, according to the principle of Archimedes, (*a*) on the brass weights; (*b*) on the material; and (*c*) what is the net correction which must be applied to the weighing on account of this effect, to arrive at the true weight, that is the weight which would be obtained in a vacuum? Take the density of the air as 0.0013 grams per cubic centimeter.

21. A cubical block of wood of density 0.6 and with sides 10 cm. long floats with its sides vertical in water. With how much force must we push down on it to get it all under water? Compare the force needed to hold it 1 cm. under with that required to hold it 10 cm. under the surface.

22. A hollow iron ball just floats in water. It contains a cavity of 10 cc. volume. How much weight of iron is there in the ball?

23. The gas bags of a large airship (the *Graf Zeppelin*) have a total volume of 100,000 cubic meters. If 1 cubic meter of air weighs 1.29 kg. and air is 14 times heavier than hydrogen gas, how much weight in all could the airship lift when filled with hydrogen?

24. In Fig. 1–7, C is a thick-walled steel cylinder, such as Bridgman uses in which to produce very high pressures. In this cylinder there is a closely fitting piston P, to which a large force F is applied from above in order to create the high pressure in the experimental chamber V. The liquid in V would leak out along the wall between C and P but a rubber washer (shown shaded in the figure) is placed above P to prevent this. Above this washer is a bar B whose lower surface is smaller in area than the piston P (E being an empty space). Show that the pressure on the washer is greater than anywhere else, so that the material in the washer tends to flow into V at high pressures rather than allowing the liquid to escape.

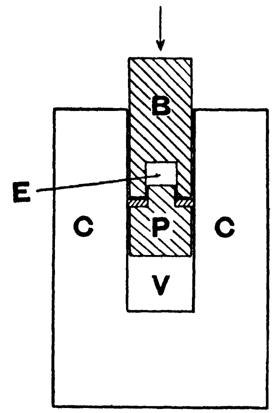


FIG. 1–7
A leak-proof piston

CHAPTER 2

AIR PRESSURE

Air density and pressure, 15; barometers, 16; barometric corrections, 16; weather maps, 17; weather predictions, 19; aneroid barometer, 19; the upper air, 19; water pumps, 21; force pumps, 23; vacuum pumps, 23; Hare's apparatus, 24.

Air density and pressure. Anyone who has been pushed about by a high wind is willing to admit that we live in a sea of heavy air. Simple experiments show that the air exerts a strong pressure due to its weight even when there is no wind to make us aware of its presence, and this pressure resembles somewhat the pressures considered in the preceding pages. Air differs from water, however, in being easily compressible. This makes the low-lying air in which we ordinarily live much denser than that found at the tops of high mountains. On this account no simple relation between pressure and depth exists in the case of air pressures such as we have found useful for liquids.

The density of the air was roughly but ingeniously measured by Galileo [1] (1637) before the days of vacuum pumps (see problem 8, p. 26). Nowadays it is easily found by pumping out the air from a large bottle, weighing the bottle full and empty, and finding its volume by the weight of water it can contain. Air under usual conditions is nearly 800 times lighter than water.

[1] Galileo Galilei (1564–1642), Italian astronomer, mathematician, and experimental physicist, perhaps the first man to whom such a title can be applied. By means of the telescopes which he constructed he was able to discover convincing evidence against the current belief that the earth was the center of the universe, with the sun revolving around it. This led him into difficulties with the Church, which were augmented by the vigor and sarcasm of his utterances in his own defence, so that he was eventually condemned and compelled to abjure his new doctrine and live in seclusion for the rest of his life. He laid the foundation of experimental mechanics, clearly formulated the laws of falling bodies (p. 46) and the first two of the laws of motion now often known as Newton's. He demonstrated from the leaning tower at Pisa the fact that all bodies fall at the same rate, using weights of 1 lb. and 100 lbs. for the test. Not the least of his achievements was his success at Padua in attracting audiences of nearly 2000 persons to his scientific lectures.

TABLE II

Density of Gases

In grams per liter (1000 cc.)
at 0° C. and 76 cm. pressure

Air	1.293
Carbon dioxide	1.98
Helium	0.178
Hydrogen	0.0899
Nitrogen	1.251
Oxygen	1.429
Steam at 100° C.	0.598

Barometers. The discovery of how to measure the pressure of the atmosphere is credited to Torricelli[1] (1643) who invented the mercury barometer. A long glass tube, sealed at the bottom and full of mercury, is inverted into a bowl of mercury (Fig. 2-1) without allowing any air bubbles to enter. If the tube is long enough, the mercury stands, not at the top of the tube, but at a definite height only, averaging about 76 centimeters at sea level, above the outer surface. Above the mercury in the tube in a properly made instrument is a vacuum, except for the quite negligible amount of mercury vapor which escapes into it from the liquid surface. Air and moisture ordinarily cling to the tube when the experiment is tried; the air may be nearly all gathered up if a large air bubble is passed back and forth over the tube while it is horizontal and nearly full, but boiling the mercury before inverting the tube is necessary to drive it all out, and this is a somewhat delicate and dangerous operation.

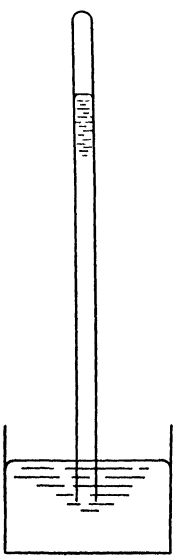


FIG. 2-1
A simple mercury barometer

Barometric corrections. Several causes make readings of the barometric height incorrect. A rise of temperature causes the mercury to expand and thus enables the air to support a slightly higher column of it. The readings are usually "corrected to zero" degrees Centigrade

[1] Evangelista Torricelli (1608–1647), Italian mathematician and physicist; disciple of Galileo. He found the law of the velocity of flow of a liquid from a hole in the containing vessel (p. 87), was the author of interesting geometrical propositions and made improvements in the microscope and telescope.

(Chapter 11). The length of the scale measuring the barometer also changes with temperature, making a slight correction necessary. If the upper surface of the mercury is not so wide as the lower, surface tension (see p. 135) lowers the narrower surface a little, and this must be allowed for. If readings of the barometer are to be compared which are taken simultaneously at different places, these places will not often be at the same height above sea level, and a special correction must be applied on this account. Tables (here omitted) are commonly supplied for use in making these corrections.

Weather maps. The United States Weather Bureau gathers data daily in regard to air pressure, temperature and other meteorological effects at over two hundred stations widely scattered over the country, as well as many received from the adjacent seas.

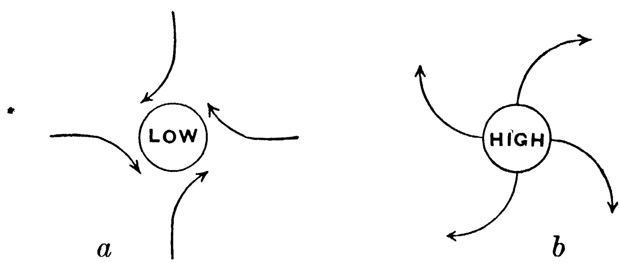


FIG. 2–2
Circulation about low- and high-pressure areas

The pressure data are corrected in the manner indicated above and then plotted as a "weather map" showing lines of equal pressures, and also lines of equal temperatures. These maps disclose definite areas, hundreds of miles wide, in which the pressure is less than the average. The winds tend to blow spirally inward toward the centers of such areas, at the same time circulating about them in a sense opposite to the motion of the hands of a clock (Fig. 2–2*a*). At the center itself there is a rising current. These low-pressure areas carry rainy conditions with them. They normally travel several hundred miles a day, moving faster the more intense they are. Their paths across North America are from west to east, southwest to northeast, or south to north along the east coast. In any case they usually leave the country at the northeast corner, and sometimes continue traveling in a northeasterly direction across the ocean.

Areas of high pressure with a spiral, outward, clockwise circulation (Fig. 2–2*b*), bringing fair and cool weather, also occur on nearly every map. These usually travel in much the same directions as the low areas. In them there is a downward current at the center bringing clear air from the upper levels. The reason for the direction of wind currents in both cases is furnished by the rightward deflection of moving bodies (in the northern hemisphere) which is explained later (p. 52). An example of weather maps is given in Fig. 2–3. It shows an area of low pressure accompanied by rainfall (in the shaded parts), and the line of arrows passing

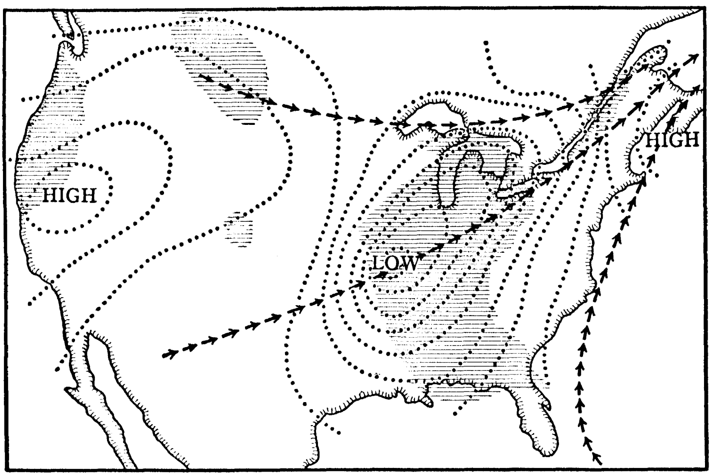


FIG. 2–3

Weather map on a day in Jan. 1928, showing typical paths of low-pressure areas; simplified by the omission of temperature lines.

through its center shows the usual path taken by such low areas. Two similar lines of arrows in the diagram show typical paths for low areas originating elsewhere. The center of this particular low area at the time shown in this figure was in Missouri; a day earlier it was in Colorado some 700 miles to the west. A day later one would have expected it to be an equal distance to the northeast, perhaps in Pennsylvania, bringing rain in the states to the southward and eastward of its center. The difficulties of weather predicting are well illustrated by the fact that in this particular case the low-pressure area grew in intensity and speed, and the center was found the next day twice as far along the path as would have been expected (i.e. just beyond Maine), leaving clear weather behind it.

Weather predictions. Abnormal behavior on the part of both high- and low-pressure areas is the bane of the weather man; they sometimes fade away without moving, or, as in the above example, travel many times faster or slower than usual, or go by an erratic path. In spite of these vagaries, weather predictions from such maps yield in skilful hands about 85% of successes. A rough general rule for the northeastern United States is that one may expect tomorrow the sort of weather existing today about 600 miles to the west or southwest.

Aneroid barometer. A portable form of barometer is the ***Aneroid,*** Fig. 2–4, in which a flat, evacuated box with an elastic cover supported by a strong external spring is connected to a multiplying device whereby motions of the center of the cover are magnified and cause a pointer to play over a scale. Evidently, an increase of pressure will drive the cover in a little, while changes of temperature will produce no effect unless there is air inside the box. The scale must be made by comparison with a mercury barometer, and often needs adjustment. The instrument may be made in watch size, and is sometimes graduated to read not pressure but feet of elevation above sea level, for the use of mountain climbers. A reading of such an instrument is taken at the foot of a mountain, again at the top, and once more at the foot on the return. Changes in the barometric pressure in the interval may make the two low-level readings disagree, but their average will usually yield accurate results. These instruments though small may at the same time be sensitive enough to indicate the height of a table above the floor. They are used in aircraft as "altimeters."

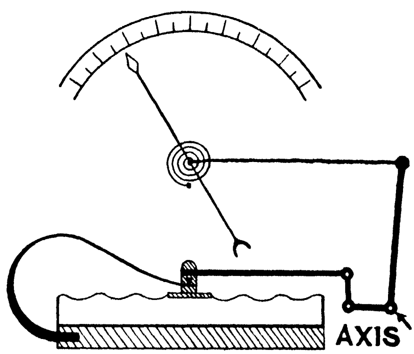


FIG. 2–4
Mechanism of the aneroid barometer

The upper air. The law of decrease of pressure with height will not be considered here, but the change is indicated by the pressure curve in Fig. 2–5. Some of the methods of study of the upper air are also indicated, as well as a few of the results. The upper regions have been studied by means of "sounding balloons,"

which carry recording instruments and bring us news of the extreme cold and the strange uniformity of conditions above a height of seven miles.

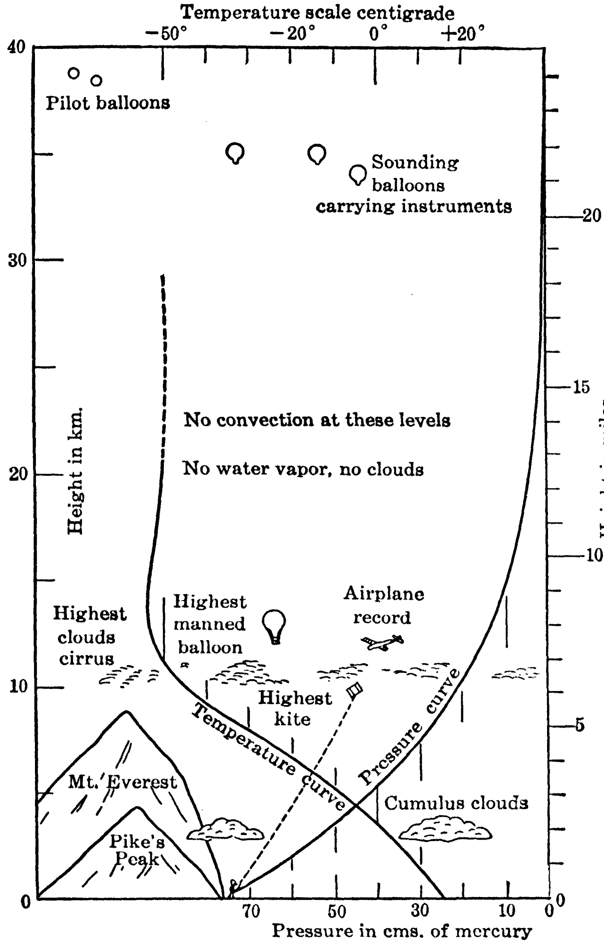


Fig. 2–5

The average pressure and temperature of the upper air in summer at latitude 45°

Example on air pressure. A man who had not studied physics planned a refrigerator on the principle of the vacuum bottle; that is, a box surrounded by a double-walled evacuated space. As a sample section of such a wall he made a box of sheet zinc, 30 by 18 inches in area, and two or three inches deep. It is interesting to calculate the force on each of the two large surfaces of this box, which presses them together when the air is exhausted from the inside. Each side has $30 \times 18 = 540$ sq. in. The usual barometric pressure is about 1033 gr. per square centimeter, or 14.7 lbs. per square inch. Hence the desired force

is $540 \times 14.7 = 7938$ lbs., or nearly four tons. One can imagine the condition of the box when the experiment was tried. It makes an instructive illustration of the magnitude of the air pressure to exhaust the air from the inside of an old tin can, provided that it is no longer needed. One such is shown in Fig. 2–6.

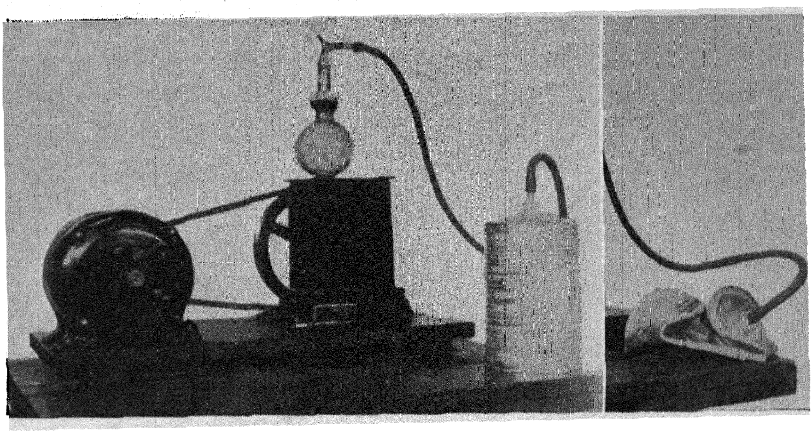


FIG. 2–6
A tin can before and after exhaustion by a vacuum pump

Water pumps. Water can be raised 25 feet or more by means of a so-called suction, or lift pump. We shall consider one type as a good, practical example of the action of air and water pressures together. Figures 2–7, *a*, *b*, *c*, illustrate the action of a simple sort of lift pump. In Fig. 2–7*a* the piston is shown going up in

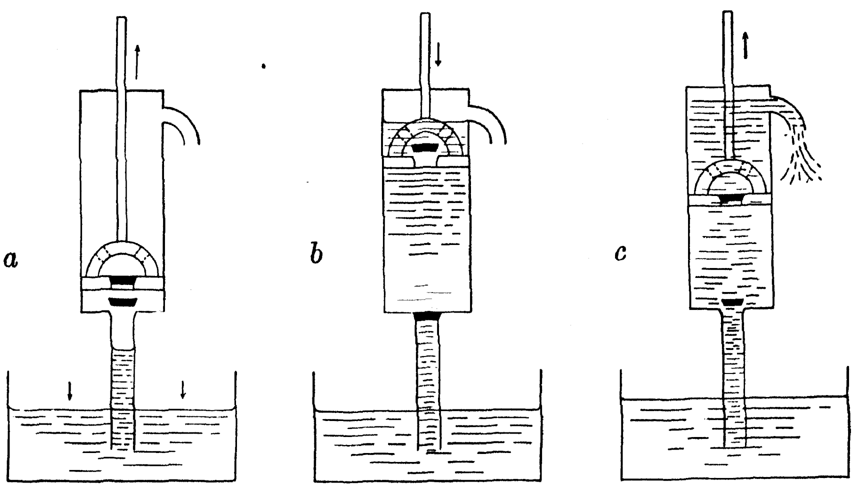


FIG. 2–7
A lift-pump

the cylinder. If it fits tightly, the air below it will be expanding, producing a partial vacuum, into which the external air pressure, acting on the surface of the water in the well below, will drive water. In passing into the cylinder the water goes by a one-way door, or valve. On the down stroke (*b*) this valve shuts, and another one in the piston itself opens, passing the water into the upper part of the cylinder, out of which the next stroke (*c*) lifts it to the spout.

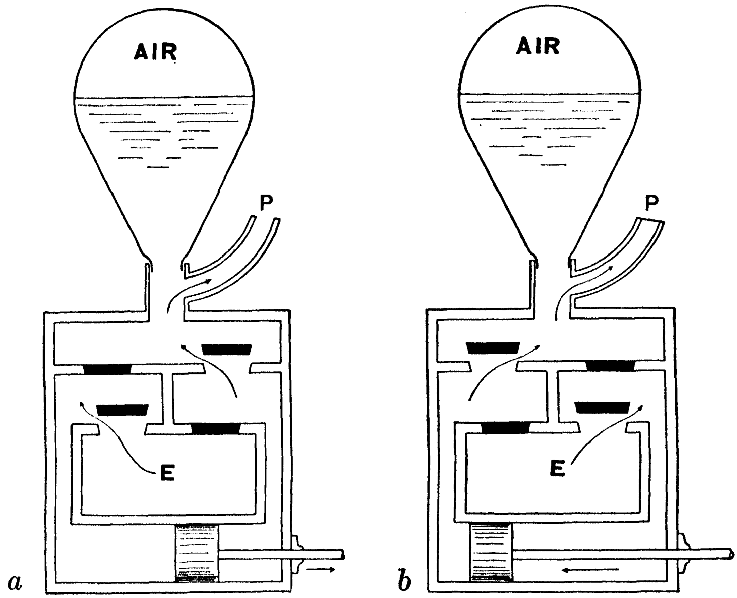


Fig. 2–8
A double-acting force pump

Galileo taught that "Nature abhors a vacuum," but it is a common experience with pumps that if the height from the cylinder to the surface of the water in the well is much over 25 feet, the upstroke of the piston creates a partial vacuum which Nature does not worry about. The average atmospheric pressure is incapable of supporting a column of mercury higher than 76 cm., or of water higher than 76×13.6 or 1033.6 cm., which is 33.9 ft. If we make allowance for the pressure produced by water vapor at the top of the column, (see p. 194) this is reduced to 33 feet; and in practice with actual pumps, which do not entirely keep the air out, a column of about 25 feet is all that can usually be lifted. It is to be noted that the action of a pump is not

properly described as "suction"; that is, in all such cases the motion that results is due to a push from behind rather than a pull in front.

Force pumps. An interesting modern form of pump is a double-acting force pump, which delivers water when the piston moves either way. Figures 2–8*a* and 2–8*b* show its action. The valves are shown in black, the springs that force them back into place being omitted from the sketch. Water enters through a pipe *E* from the well to a chamber from which it passes through a first valve into one side of the piston chamber, out of which it

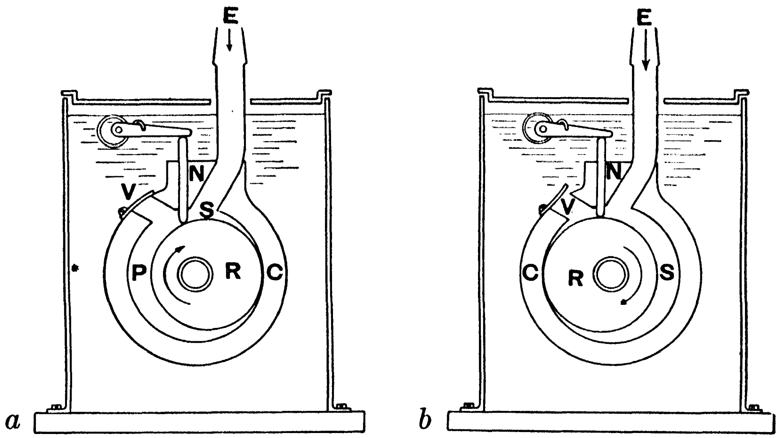


FIG. 2–9

A rotary oil vacuum pump. The complete machine is seen in action in Fig. 2–6.

is later driven by the return of the piston through an outlet valve into the top of the pump. At *P* is connected a system of piping which may carry the water to a great elevation, if the power driving the pump is sufficient. An air chamber is supplied so that the pressure of the compressed air in it will always be acting and will tend to make the flow steady.

Another form of pump in common use, the "centrifugal pump," is considered later (see p. 72).

Vacuum pumps. The usual form of air pump for producing a vacuum involves a cylinder and piston with valves, not very different in plan from water pumps. The valves must be very light, and in the better forms are sealed with oil. A more modern and convenient design is a rotary air pump whose operation will be clear from Figs. 2–9*a* and 2–9*b*. *R* is an eccentric rotor, ro-

tating inside a closed cylinder, and making close contact with its wall at one point C, (or rather along a line at C, since the apparatus has a considerable thickness, perpendicular to the plane of the paper and the diagram gives a plane section only). N is a vane or barrier which slides up and down as the rotor passes around, and is always held tightly against the surface of the rotor by a spring. As the rotor turns (from the position in (a) to that in (b)) it enlarges the space S into which the air enters from the vessel to be exhausted, connected at E. At the same time, the space P is diminishing and the air in it is finally driven out past a light escape valve V, sealed with oil. The passage of the contact point C by the top of the cylinder cuts off another quantity of air, and drives it out in its turn. In a popular form of this pump, two of these devices are connected in tandem, so that the one exhausting the vessel at E drives out its portion of air, collected in each complete turn, into a space which is itself exhausted by the second pump. In this way the pressure needed to open the first escape valve is much reduced, and the combined action of the two pumps is capable of producing a pressure as low as 0.001 mm. on a mercury gauge, which is less than one-five hundred thousandth of the pressure of the atmosphere.

Another form of pump for obtaining the highest vacua is discussed later (p. 141).

Hare's apparatus. An interesting experiment involving pressure is furnished by the apparatus shown in Fig. 2–10 for comparing the densities of two liquids. The liquids are in vessels A and B into which glass tubes dip, connected together above, with an opening E at the top, controlled by a stopcock. Air may be sucked out of E so that the liquids rise in their separate tubes.

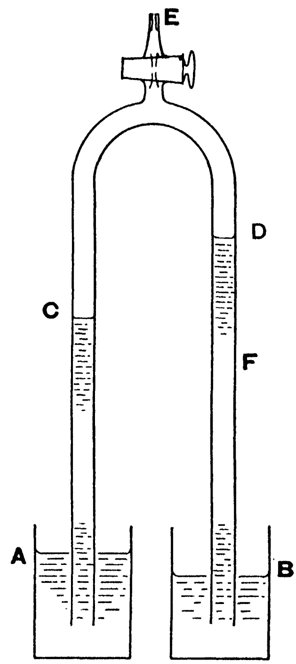


FIG. 2–10
Hare's apparatus

In the air space above C and D the pressure is the same throughout, and is a certain amount less than that outside. At the open surfaces of A and B the pressure is equal to that of the atmosphere. Therefore, in ascending from the level of the surface at A inside the tube to C, or likewise from B to D, the same change of pressure is encountered. If h_1 and h_2 are the heights and

d_1 and d_2 the densities of the liquids, on one side the pressure change is measured by h_1d_1 and on the other by h_2d_2. Hence, $h_1d_1 = h_2d_2$, or $\frac{d_1}{d_2} = \frac{h_2}{h_1}$. The densities are thus in the inverse ratio of the heights. This makes a very simple method of measuring liquid densities with moderate precision.

Example on air pressures. Men digging a tunnel under a river work in air-tight chambers or caissons, in which air pressure is maintained by a suitable pump at so high an amount that, if a break occurs, air will rush out rather than water pour in. What air pressure, as measured by a barometer inside the working chamber, will be sufficient if the work is done 15 m. below the surface of the water, on a day when the barometer outside stands at 77 cm.?

The pressure due to 15 m. of water is 1500 gr./cm.^{2} Pascal's principle reminds us that the ordinary pressure due to the atmosphere is also felt in the caisson. That is, 77×13.6, or 1047 gr./cm.^{2} must be added. The total pressure is therefore 2547 gr./cm.^{2} This would hold up a column of mercury 2547/13.6 or 187.3 cm. high.

PROBLEMS

1. If the fluctuations of a mercury barometer with the weather are from 78 to 74 cm. height, how far would a barometer made of water change at the same time, assuming that the space above the water surface inside remains always a perfect vacuum?

2. Find the pressure in grams per square centimeter on the air in a diver's suit, 30 m. down in sea water of specific gravity 1.03, the atmospheric pressure being 1000 gr. per square centimeter at the time. It is understood that the air is kept bubbling out of the suit through a valve by means of the pressure within.

3. In modern water systems for country houses a good pump drives water into the bottom of a large steel tank from below which contains air. The tank (Fig. 2–11) is located in the cellar, but the pressure of the air is sufficient to force water all over the house even when the pump is not running. If the air pressure is 60 lb./in.^{2}, find the pressure at a faucet 30 ft. above the surface of the water in the tank. (N.B. The pressure in such cases, as with automobile tires also, is the excess pressure over that of the atmosphere.) In this arrangement, if it were possible to put a suitable barometer inside the air in the tank, what height would it read?

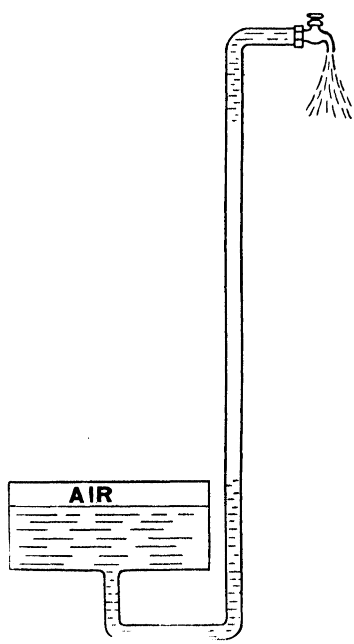


FIG. 2–11
The air in the tank in the basement forces the water upstairs.

4. Find the total force in tons, due to atmospheric pressure, on the surface of a man's body, assuming his total outside surface area to be 10 sq. ft. How is it that he is not crushed to death by this force?

5. If a tiny hole were opened in one of the tubes of Fig. 2–10 say at F, would the liquid begin to pour out, or would air first bubble in?

6. Would it make any difference in the results with Hare's apparatus (*a*) if the bottom of the left-hand glass tube almost touched the bottom of the vessel A, while the right-hand tube ended only a little below the surface of B? (*b*) if the left-hand tube were twice the diameter of the right-hand one? (*c*) if one tube were conical, larger at the top than at the bottom?

7. What is the average density of the air if, when a man climbs three flights of stairs (12 m.), a barometer which he carries with him shows a fall of pressure from 76.00 cm. to 75.89 cm.?

8. Galileo weighed a bottle full of air only. Then he corked it up tightly and forced water into it until the air was compressed into about a fourth of the volume it previously occupied. The bottle was again weighed in this condition. Then the air was allowed to escape (as much as it wanted to), and the bottle was weighed a third time. Show how the weight (and density) of the air can be obtained in this way.

9. What would be the reading of an aneroid barometer, if it could be placed inside an automobile tire whose air pressure as read by a tire gauge is 30 lbs. per square inch? (N.B. The gauge reads excess of pressure within over that without.)

CHAPTER 3

STATICS. SOLIDS AT REST

We are constantly surrounded in our homes and cities by the evidences of man's skill in mechanical operations. Even in times long ago it was noteworthy. The Egyptians must have understood how to lift and handle huge blocks of stone, or the pyramids could never have been built. In the Middle Ages much more had been learned. Leonardo da Vinci,[1] though best known to us as a painter, was a very inventive engineer and his notebooks show that he was acquainted with a quite modern range of mechanical principles. In our own observations we see wide rivers being spanned by bridges, or heavy weights raised by lofty cranes, and most of us are inclined to wonder a little whether the work will stand or collapse. It is interesting to see how to find out just what forces are involved in any such operation and just how strong any part of a structure needs to be in order to do what is expected of it. The principles by means of which the forces can be found are easy to understand, and we shall consider a few simple examples.

In very many practical cases the parts of a structure are at odd angles to each other, and we are first faced with the necessity of handling forces in different directions. For this the operations known as the ***composition*** and ***resolution*** of forces are a great convenience.

[1] Leonardo da Vinci (1452–1519); Italian painter, architect, sculptor, engineer, and (for his day) scientific man. He spent much time and thought on mechanical methods of attack and defense. There is an excellent account of one aspect of this many-sided genius in "The Mechanical Investigations of Leonardo da Vinci" by Ivor B. Hart, 1925 (Chapman and Hall).

Vectors. A force is an example of a ***vector quantity.*** By this is meant a quantity that has magnitude and direction and can therefore be satisfactorily represented by a line of definite length. This line must be drawn in a direction parallel to that of the force, and of a length proportional to the amount of the force, in convenient units. Other examples of vector quantities are velocity, momentum, etc. Quantities that have magnitude only, without any sense of direction, are called ***scalar*** quantities; examples are volume, density, temperature, etc.

Composition of forces. All vector quantities can be combined by the ***parallelogram rule,*** a process we shall have occasion to use very frequently. As a simple example, if two persons pull with cords upon a small object O, Fig. 3-1, the forces along the cords might be represented by OA and OB. The rule then says that the ***resultant*** OC, the diagonal of the parallelogram formed with OA and OB as sides, is equivalent in every way in its action on O to the forces OA and OB combined. In fact, OA and OB may be done away with, and a pull represented in direction and amount by OC substituted for them. This very reasonable statement can be shown to hold true in general. It may be tested experimentally by setting up spring balances to pull in such directions as OA and OB and then applying a third force equal and opposite to OC, which will then be found to annul the action of OA and OB together.

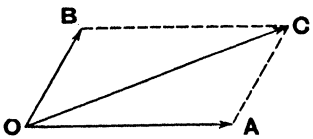


Fig. 3-1

Resolution of forces. Conversely, a pull OC can be ***resolved into components*** OB and OA along any two directions, which may then take its place. This process of resolution may be carried through for directions at any angle to one another, but is more frequently done along directions at right angles, as in Fig. 3-2a. These are often horizontal and vertical, as in the figure, but may be in other directions. A general rule for finding the component of a force OC [Figs. 3-2b and 3-2c] in any direction, say OF, is to drop a perpendicular from C on OF (produced if necessary) which cuts off on OF a line OB. This is the required component. The

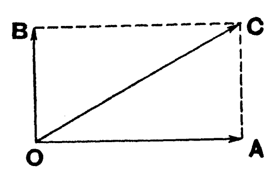


Fig. 3-2a

component is sometimes called the "effective part" of the force in the chosen direction. If the object O were constrained so that it could move only in the line of OF, we could say that one part of OC, namely OB, produced the motion while the other com-

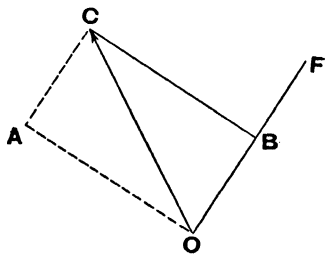


FIG. 3–2*b*

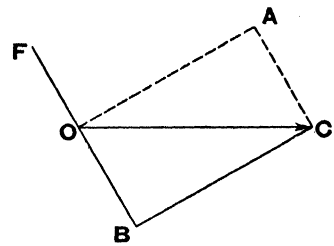


FIG. 3–2*c*

ponent OA in the rectangular figure was ineffective and could not help, since it was at right angles to the only possible motion.

Calculation of component. If we wish to calculate (rather than draw) the size of the component of a vector quantity which makes an angle θ with the vector, all we have to do is to multiply the vector by the cosine [1] of the angle θ, (assuming as usual that the two components are to be taken at right angles to each other). Thus in Fig. 3–2*b* the angle the vector makes with the chosen direction is COF and the cosine of this angle is OB/OC, which when multiplied by OC yields OB, the desired component.

Geometrical addition. If OA and OB can be combined into OC, the process is sometimes called ***geometrical addition,*** and to save time we may occasionally write $OA + OB = OC$, remember-

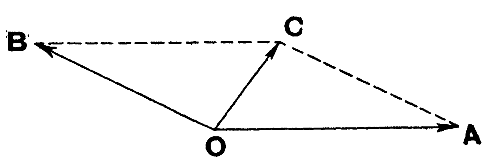


FIG. 3–3

ing, however, that in this sort of addition we may find the geometrical sum much smaller than either of the two vectors added. For instance, in Fig. 3–3 OC is small compared with either OA or OB, because the directions are such that OA and OB nearly annul one another.

[1] See the appendix for trigonometrical symbols and their use.

Forces in equilibrium. If a body is at rest and forces are acting on it, they must be so arranged that their net effect is zero. If there are only ***two*** forces these must be equal and opposite. If there are ***three*** and they are in different directions, the resultant of any two must be equal and opposite to the third. This brings us to the useful rule that ***three forces which balance must meet in a common point***, (unless they are parallel).

We may represent three balanced forces in another way. If they are shown as OA, OB and OC in Fig. 3–4*a*, they may also be drawn as in Fig. 3–4*b*. The latter figure is constructed by drawing OA first, and from A drawing a line AF representing OB, that is,

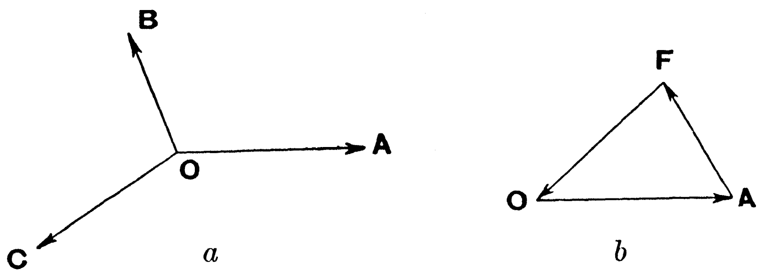


FIG. 3–4

equal and parallel to it; then the line FO is drawn, completing the triangle. This triangle is half of such a parallelogram as is shown in Fig. 3–1 (p. 28). Hence by the parallelogram rule, OF must be the resultant of OA and AF, or, FO is the force which neutralizes the combined effect of OA and AF. Hence the parallelogram rule leads to the conclusion that three (non-parallel) forces producing equilibrium can be drawn as the sides of a triangle.

We may extend this method of representing forces in equilibrium; if there were four of them forming such a group, they would, when drawn in this manner, form a closed four-sided figure. In general, any number of forces in equilibrium form a closed polygon, and this method of representing them is usually known as the ***polygon method.***

Practical examples. It will be of interest to illustrate these ideas by considering a few practical cases, and using the foregoing rules to derive results.

(1) A garage truck is towing a disabled automobile up a hill whose inclination is such that there is 1 ft. vertical rise for each 9 ft. along the slope (Fig. 3–5).

The automobile weighs 2700 lbs. Find the force in the tow-rope which is needed to hold the car at rest on this slope.

The car is acted on by its own weight, vertically downward, by a force P perpendicular to the slope due to the roadbed pressing against the wheels, and by a third force, that in the tow-rope. These three forces must meet in a common point. The resultant of any two of these forces must be equal and opposite to the third; in particular, P and F must form sides of a tilted rec-

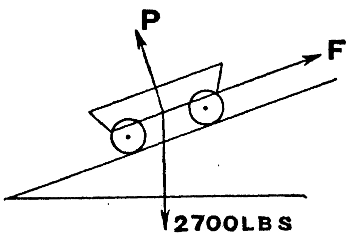


FIG. 3–5

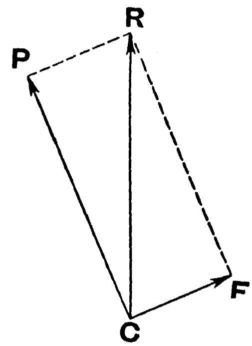


FIG. 3–6

tangle, whose diagonal, their resultant, is equal and opposite to the weight, 2700 lbs. From the similarity of the triangles in the force diagram Fig. 3–6 and in Fig. 3–5 it follows that the ratio of the rope force CF to the diagonal CR must be the same as the ratio of the vertical height of the hill to the slope length. Hence

$$\frac{CF}{CR} = \frac{\text{height}}{\text{slope}} = \frac{1}{9},$$

or
$$CF = 2700/9 = 300 \text{ lbs.}$$

(2) A 150-lb. man, W, is hanging by his hands at O to a light rope ROS, the parts of which make angles with the horizontal RS as shown in Fig. 3–7.

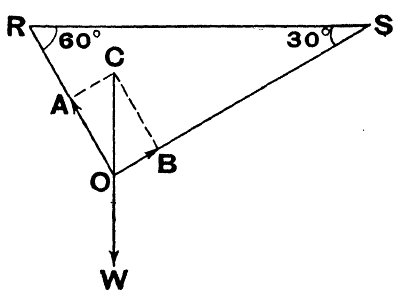


FIG. 3–7

Find the forces in the two parts of the rope. A line OW is drawn vertically down from O, its length representing on some convenient scale the weight, 150 lbs. Then a line OC is drawn, equal and opposite to OW. This must be the resultant

of the forces OA and OB in the two parts of the rope; for, if not, the man could not be in equilibrium. From C we may draw lines parallel to OS and OR, the two parts of the rope, which will cut off on the line OR the force OA, and on the line OS the force OB. Then we can measure these off to scale and the problem is solved, at least approximately

The first principle of statics. We may solve any such problem as the last by calculation rather than by a drawing, and with a great deal more accuracy. We use in such cases a piece of common sense which is known as ***the first principle of statics; the sum of the components of a set of balanced forces in any direction must be zero.***

In the last problem the vertical forces are $OA \times \cos 30°$ and $OB \times \cos 60°$ upwards, and OW itself downwards. Hence, using this principle for vertical forces,

$$OW = OA \times \cos 30° + OB \times \cos 60°$$

or

$$150 = OA \times (0.866) + OB\ (0.5) \qquad \ldots\ldots\ldots\ldots \quad (a)$$

The horizontal components are $OA \times \cos 60°$ leftwards and $OB \times \cos 30°$ rightwards; OW has none. Hence

$$OA \times \cos 60° = OB \times \cos 30°$$

or

$$OA \times (0.5) = OB \times (0.866) \ldots\ldots\ldots\ldots \quad (b)$$

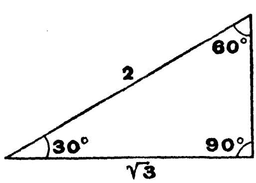


Fig. 3–8

Substituting in equation (a) the value of OA found from (b) we get an equation for OB which yields $OB = 75$ lbs.; and then (b) gives us $OA = 129.9$ lbs.

For those who prefer to avoid the use of trigonometry, it will be convenient for some problems to remember the ratios of the sides of a 30°, 60°, 90° triangle, as shown in Fig. 3–8. We could have used these numbers instead of the trigonometrical expressions above; e.g. $\cos 30° = \frac{\sqrt{3}}{2}\cdot$ In practise, however, we seldom run into such simple cases; with other angles occurring, it is necessary to turn to trigonometry for help.

Moment of force. Torque. If a body is acted on by ***parallel forces*** applied at ***different*** points, and is nevertheless at rest, the sum of the forces in one direction must equal the sum in the op-

posite direction, in accord with the first principle of statics. But here another effect also must be considered. Since no turning can be allowed to occur, the tendency of each force to produce turning must be exactly annulled by the opposite tendency of all the other forces. In thinking of this, we may choose ***any*** point as that about which we imagine the turning to be possible. The tendency of a force to produce turning is called the ***moment of the force,*** or the ***torque.*** It is equal to the force itself multiplied by the shortest distance from the axis of rotation (the fulcrum) to the line of the force. This distance is sometimes called the lever arm of the force. Thus, if anyone opens a door (Fig. 3–9) he pulls on the handle H usually in a direction perpendicular to the door, because thereby he gets the greatest torque, as defined above. If AH is the door, with the axis of rotation at the hinge A, and if one pulls in the direction HF with a force F, the lever arm is p, and the moment of force or torque is Fp, which is evidently less than $F \times AH$, the moment of force for the usual direction of pull. By pulling in the direction AH no torque at all would be produced, and of course the door would not open.

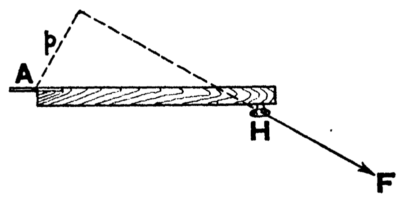


FIG. 3–9

Couples. Turning always takes place because of the action of a ***couple.*** This is defined as a pair of equal, parallel and opposite but not coincident forces. The torque produced by the couple in Fig. 3–10 would be Fl, where l is the perpendicular distance between the forces; and this same result would be obtained whatever point we chose as an axis of rotation. In the case of the door in Fig. 3–9 only one force is shown as acting, and yet there must be another force to form a couple. In this case it is hiding; the wall is exerting it on the door through the hinge. This is shown by the fact that if the hinge broke, the door would fall forward instead of turning.

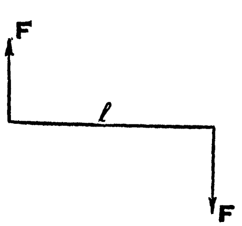


FIG. 3–10
A couple

The second principle of statics. A torque tending to turn a body in a sense opposite to the hands of a clock is arbitrarily chosen as positive. A formal statement may now be made of the

second principle of statics; namely, ***the sum of the torques acting on a body, taken about any point, must be zero,*** if the body is to remain at rest; or ***the sum of the positive torques must be equal to the sum of the negative ones.***

Levers. As the simplest illustration, the action of a ***lever*** may be considered. The common crowbar will do. If AB, Fig. 3–11a, represents a crowbar, whose weight we neglect, F the fulcrum

FIG. 3–11a

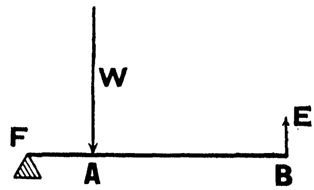


FIG. 3–11b

about which it may turn, W a heavy weight to be lifted, and E the force exerted at the end B, the bar will be balanced if the torques produced by the two forces about F are equal and opposite; or if $W \times AF = E \times FB$.

Sometimes the lever is arranged with the fulcrum at one end. As before, the lever (Fig. 3–11b) balances if $W \times AF = E \times FB$; E in this case must be directed upwards.

The forces in balanced levers are not necessarily parallel. The following is a common example. In the case of a hammer used to pull a nail out of a wall the large force W (Fig. 3–12) which is exerted by the nail on the hammer is perpendicular to the line (drawn dotted) joining N, the head of the nail, to F, the fulcrum at the point of contact with the wall. The force E should be exerted perpendicularly to the other dotted line HF. On pulling the nail, $W \times NF = E \times HF$. The forces are marked as acting ***on*** the hammer. It would not be correct to draw the force E acting on the hammer, and the force W exerted through the hammer and acting on the nail. We must, as we might say, "isolate" the hammer, and consider

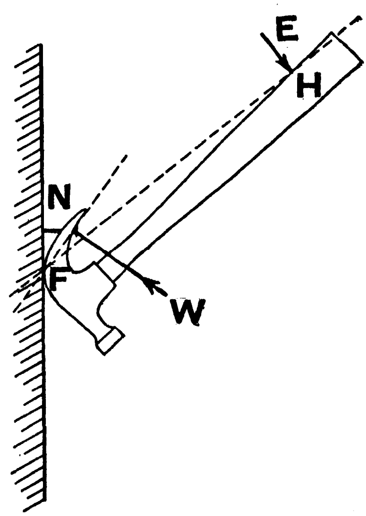


FIG. 3–12

all the forces as acting on it. Then we see that the torques must be equal and opposite for equilibrium, even though the forces are in unrelated directions.

Isolating a body. It is a great help in problems of the sort just treated to consider a body separately, and to draw the forces acting on it as though it were isolated. All forces have a double nature, acting in opposite directions at once (p. 67); thus, as a man sits in a chair his body pushes on the chair and the chair pushes back with an equal force on his body. If the force is to be represented as a vector, its direction must be specified by an arrow. The direction of this arrow in any case is made obvious by the process of isolation.

Consider an example. Suppose that a hinged box cover C (Fig. 3–13) is held up by a prop. We may isolate the cover, drawing a dotted line about it, if we prefer, to indicate its isolation. In this condition we see clearly that the cover is acted on by its weight W, by the force F exerted *by* the prop *on* the cover, and a force H exerted *by* the hinge *on* the cover. If we had been interested in the prop we should have isolated it, drawn F in the opposite direction, put in the weight of the prop, and a force at its lower end acting on the prop from its support. In any case the mental process of imagining the body to be isolated helps in resolving any doubts as to the proper direction of the forces.

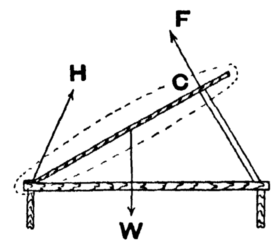


Fig. 3–13

Example with parallel forces. The following is a very practical case. AB (Fig. 3–14) represents a uniform pole, 10 ft. long, carried on its ends by the shoulders of two men. The pole weighs 20 lbs. and a 60-lb. weight is hung on the pole at a point C, which is 3 ft. from A; to find the weight borne by each man.

The weight of the pole can be taken as acting through its ***"center of gravity"*** [1] which is at the middle of the pole, since it is uniform. Evidently the men's shoulders exert upward forces P and Q on the pole. The first principle of statics tells us that

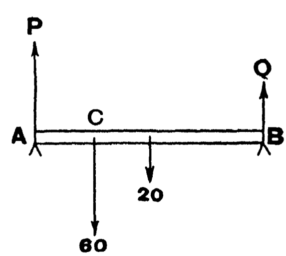


Fig. 3–14

[1] The center of gravity is the point at which the pull of the earth on the different parts of the body may be regarded as concentrated, as shown by the fact that the body will balance if supported at that point only.

the sum of the upward forces is equal to the sum of the downward ones; or

$$P + Q = 60 + 20 = 80 \text{ lbs.}$$

The second principle tells us that no turning must occur, and we may choose any point, say B, around which we may consider turning likely to happen. The torque produced by P is then $P \times 10$, and, being clockwise, is negative. The torque produced by the 60-lb. weight is $7 \times 60 = 420$, and is positive. The torque of the weight of the pole is $5 \times 20 = 100$, positive. The torque of Q is zero, since it acts through the chosen fulcrum B. Hence $420 + 100 = 10P$, or $P = 52$; whence, $Q = 80 - 52 = 28$.

Example with forces in a plane but not parallel. It often happen that several forces acting on a body are in the same plane, but are not parallel. In such cases, as before, the two principles of statics lead directly to a solution. Consider this case.

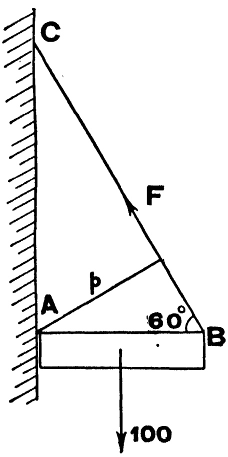


FIG. 3–15

A street sign, Fig. 3–15, is 8 ft. long and weighs 100 lbs. It is supported by an iron pin at A, which rests in a cavity in the wall and by a chain BC running up to the wall at an angle of 60° as shown. Find the force F in the chain, and the force R (not drawn) between the wall and the pin at A.

It is convenient to begin by using the second principle of statics, and considering A as the place about which turning may occur, for the reason that this gives us an equation which does not contain the rather troublesome force acting through A, which is unknown not only in amount but in direction also. To find the torque produced by the force F about A we need the perpendicular distance p. This is one side of a 30°, 60° triangle, with AB as its greatest side. Hence $p = AB \times \cos 30° = 8 \times (0.866) = 6.93$ ft.; so that the torque due to F is $6.93F$, and is positive. The torque due to the weight is 4×100, negative. Hence $6.93F = 400$, or $F = 57.7$ lbs.

To get the force at A, consider vertical forces, using the first principle. The vertical component of F is $57.7 \times \cos 30° = 50.0$ lbs. The weight 100 is all vertical. The difference $100 - 50 = 50$ lbs. must be borne by the support at A. But there is also a horizontal

force at A. Since F has a horizontal component, which is 57.7 $\times$ cos 60 = 28.8 lbs., leftward, and the weight has none, being all vertical, it follows that the horizontal force at A must be 28.8 lbs., rightward. Hence the total force at A is given by the resultant of 28.8 lbs. to the right, and 50.0 lbs vertically up. This can be found graphically, or by calculation (the square of the hypotenuse is equal to the sum of the squares of the sides), to be 57.7 lbs. It acts in a direction given by the angle whose tangent is 50/28.8 = 1.732, which is 60° from the horizontal. We could have arrived at this result more simply. Since the force F has as components 50 lbs. vertically up and 28.8 lbs. leftward, and the force at A has as components 50 lbs. vertically and 28.8 lbs. rightward, therefore the two forces are equal and equally inclined, the one inward, the other outward.

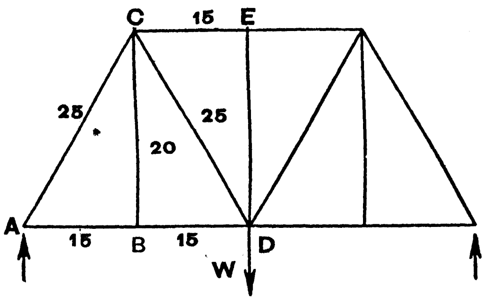


FIG. 3–16
The bridge truss

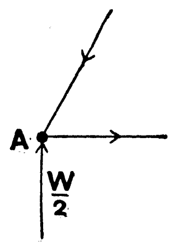


FIG. 3–17

The bridge truss. An example of unusual interest is furnished by the bridge truss, which we shall next consider, in a much simplified form.

Suppose that Fig. 3–16 represents a bridge whose weight is small compared with the load W placed at its middle point. The dimensions of the bridge are shown in the figure in feet. The bridge is supported by piers at its ends. We may "isolate" the point A, as in Fig. 3–17, and consider all the forces acting on it, in spite of which it remains at rest. Since the load W is borne equally by both piers, there must be an upward thrust at A from its pier equal to $W/2$. To balance this there must be a downward thrust in CA of such a size that its vertical component is $W/2$. There can be no vertical force in the horizontal bar or "member" AB, since all the members of the bridge are supposed to be pinned together, and to be able to push or pull along their lengths only.

Since CA is pushing on A, in return A must be pushing on it, and this means that CA is in a state of compression. CA must be pushing A to the left, as well as downwards, and there must be a state of tension in AB, pulling A as strongly to the right as the force in CA is pushing it leftward. As the triangles in this simple bridge are all "3, 4, 5" triangles, the force in CA that will have $W/2$ as a vertical component must be $5/4 \times W/2 = 5W/8$. This will create a horizontal component of $3W/8$, which is therefore the tension in AB.

One may next proceed to isolate the point B and show that there can be no force at all in CB, for, if there were, it could not be balanced, and the point B could not be at rest. Turning to the point C, we may consider vertical forces, and hence find the force in the member CD; then horizontal forces yield the force in CE; and so forth. It is thus possible to discover the force in each member of the bridge and to determine which members are in a state of compression, and which in tension.

It is interesting to shift the load, say to the point B, and re-examine the forces. The bridge is then not symmetrical; certain members formerly in a state of compression are now in tension, and one member that was idle now has a great deal to do. This part of the problem is left to the reader. It should be remarked that the calculation of the forces in actual bridges is much more complicated, for in them the weight of the bridge itself is no longer negligible.

More than three forces acting in a plane. In cases where there are more than three forces acting on a body, all in one plane, they do not as a rule any longer pass through a common point, but the principles of statics can be used to find them as before. As an example consider the following problem. Imagine a painter whose weight is 150 lbs. standing on the outer end of the street sign treated in the problem of Fig. 3–15. He adds a fourth force, a vertical one, but not coincident with the weight of the sign. To solve the problem, first find the torques produced by all the forces, as before, about the point A. We obtain $6.93F$ as the torque (positive) produced by the chain force; 8×150 (negative) as the man's torque, and 4×100 as the torque due to the sign. Hence, $6.93F = 1200 + 400 = 1600$, or $F = 231$ lbs. We easily find the vertical component of F to be 200 lbs., and its horizontal component 115.5 lbs., which must also be the horizontal component

of the force at A. The vertical component of the force at A is the total downward weight, 250 lbs., less the vertical part of F, 200 lbs., or 50 lbs. net. The total force at A is now different from what it was, and quite unlike the force F.

Forces not all in one plane. Still more complicated problems are commonly met with, in particular those in which a body is acted on by several forces not all in one plane; but these are left for more advanced study.

Hints on the solution of problems in statics. It is impossible to give directions for the solution of all sorts of problems. It is the very essence of a good problem that it offers something new, and must be thought out by itself. In general, however, the following hints may be offered:

(*a*) Isolate the body (or the parts of the body one at a time) to which the problem applies, and thus settle the directions of the forces.
(*b*) Locate the points of action of the forces, known and unknown. Represent these by definite straight arrows when both magnitude and direction are known, and by curly arrows if they are unknown.
(*c*) If the forces are parallel apply the first principle of statics and then the second.
(*d*) If all the forces act through a point try the polygon method (p. 30); or
(*e*) Take components of the forces along certain directions, and then apply the first principle to these; or
(*f*) Apply the second principle, considering torques about an assumed center of rotation. Choose such a point as will make the least known force disappear from the calculations; i.e., a point through which it acts.

Whether to try (*d*), (*e*), or (*f*) first will depend on the problem. Often one can see with little difficulty that one of these will yield an immediate result and thus let in much light on the rest of the problem. But the only real way of learning how to solve a new problem is to have solved a great many others.

Stable and unstable equilibrium. As we have seen, a body acted on by several forces can still be at rest. If the body is displaced a little, the forces may make it return to its original position, in which case it is said to be in ***stable*** equilibrium; or it may remain in the new position with no tendency to move either way, when it is said to be ***neutral;*** or, finally, it may continue moving until it reaches some quite different position, in which case it must have been in ***unstable*** equilibrium.

An egg resting on a table is in neutral equilibrium with reference to a displacement which rolls it along a level table, but in stable equilibrium when tipped a little in the direction of making it stand on end. If balanced on its tip it is in unstable equilibrium.

An ordinary chair is in stable equilibrium if tipped a little, but if tipped more and more it finally reaches a position of unstable equilibrium.

What determines the sort of equilibrium is the behavior of the center of gravity of the body when it is moved. If this is raised by a small displacement, the equilibrium is stable because the center of gravity tends to fall back; if it is unchanged, the equilibrium is neutral; if lowered, the position is unstable. Any body tends toward the stable position in which ***its center of gravity is as low as possible.*** In terms of energy (Chapter 6) a more general statement may be made, which is that any body (or connected system of bodies) tends toward that position in which its potential energy is as low as possible.

PROBLEMS

1. Devise an arrangement of at least three forces acting on a body (not applied at the same point) which (*a*) satisfies the first principle of statics but not the second; (*b*) satisfies the second (if a particular point is chosen as fulcrum) but not the first.

2. A rope supports one end of a pole, the other end of which rests on a smooth level ice surface. What direction will the rope take?

3. A 10-lb. weight hangs by a rope. If a horizontal force holds it out at an angle of 30° to the vertical, how large is the force and how great is the tension in the rope?

4. The two ropes of a swing each make an angle of 30° with the vertical. Together they support a 50-lb. child, at rest. Find the tension in each rope.

5. A man carries a uniform pole 6 ft. long and weighing 10 lbs. over his shoulder, holding it horizontally. His hand pulls down on one end; his shoulder supports the pole at a point 2 ft. from the same end. A 5-lb. bundle hangs from a point 1 ft. from the other end of the pole. Find the force exerted by the man's hand, and the force against his shoulder.

6. A man is lifting one end of a plank, so that the plank makes an angle of 30° with the ground. The plank is uniform, 12 ft. long, and weighs 100 lbs. Assuming that the man's force is at right angles to the plank, find how much he must exert to hold the plank in this position. Show an isolation diagram of the plank.

7. A 150-lb. man is standing on a horizontal shed roof, pulling up a 60-lb. box from the ground by means of a rope. The rope is vertical between the box and the edge of the roof, and passes over the edge (assumed to be frictionless) toward the man's hand, making then an angle of 30° with the horizontal. Find the force exerted by the man to hold the rope, and also the total force between the man's feet and the shed roof (the latter graphically, if preferred).

8. A painter's staging consists of a narrow platform 12 ft. long weighing 100 lbs. and hung by a vertical rope at each end. Two 150-lb. men are standing on the platform, one in the middle, the other 2 ft. from one end. Find the forces in the two ropes.

9. A builder makes a horizontal staging out of a 100-lb. plank 16 ft. long supported symmetrically on two brackets 12 ft. apart, so that the plank projects 2 ft. at each end beyond the supports. A 120-lb. man stands on one end of the plank. Find the forces on the two supports.

10. A light rope 100 ft. long is fastened loosely between two equally high supports, and a weight of 100 lbs. is attached to the middle of it. The middle of the rope then stands 20 ft. below the level of the supports. Find the tension in the rope.

11. A street car trolley 10 ft. long weighs 20 lbs. and is inclined at an angle of 45° to the horizontal. A light rope is attached to the upper end of the trolley, making an angle of 60° with the trolley. A pull of 20 lbs. on the rope will just release the trolley from the wire along the under side of which it runs. Find the least torque, applied by the spring at the base of the trolley, which is needed to keep it in position. Find also the force (all vertical) with which the trolley presses against the wire.

12. A man holds a 10-lb. weight in his hand, keeping his forearm at rest in a horizontal position. In Fig. 3–18 E is supposed to be his elbow and H his hand, while F is the force in the muscle attached at M, which is assumed to make an angle of 60° with the horizontal. If EM = 3 in., and EH = 15 in., what will the force F be, and also the force acting at E which is necessary to produce equilibrium?

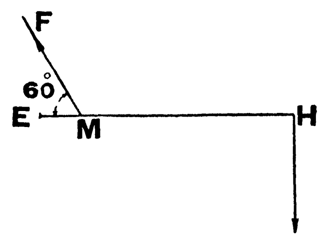


Fig. 3–18

13. A uniform ladder 10 ft. long weighs 50 lbs. It rests against a smooth vertical wall at an angle of 30° to the wall. A 150-lb. man is standing on the ladder 3 ft. up from the base. Make an isolation diagram and find the frictional force at the base necessary to keep the ladder from slipping. (N.B. If the wall is smooth, the force against it must be perpendicular.)

14. A ladder whose weight is 90 lbs. and length 10 ft. leans against a smooth vertical wall, making an angle of 30° with it. The center of gravity of the ladder is 4 ft. from the bottom. A man standing out in front of it grasps it at its middle point and pulls horizontally just strongly enough to bring the top of the ladder away from the wall a small fraction of an inch. How much force is the man exerting?

15. An automobile is stuck in a mudhole. To get it started the driver ties a strong rope to it and to a large tree 40 ft. ahead, pulling the rope as tight and straight as possible. He then pushes sideways on the rope at its middle point, and by exerting a force of 100 lbs. he moves that point a distance of 2 ft. As-

suming the rope not to stretch, how much tension is there in it then, pulling on the automobile?

16. Boards are being passed up from a wagon into the second-story window of a new building. Each is 16 ft. long, uniform and weighs 50 lbs. The man at the window has hold of one by its upper end, and a point 3 ft. from that end is in contact with the edge of the window sill. The board then slopes down, making an angle of 30° with the horizontal. Find the force the man must exert to keep it in this position, assuming that the window sill is frictionless, and that the board does not bend.

17. A gate 6 ft. long and 4 ft. high, weighing 100 lbs., is hung on two hinges 3 ft. apart vertically. The hinges are so arranged that half the weight is borne by each hinge. Find the total force on the gate at each hinge, and the angle at which it acts.

18. The four legs of a folding table form a rectangle on the floor 3 ft. $\times$ 1 ft. It has two drop leaves, each weighing 10 lbs. One is hanging down and its weight then acts through a point 4 in. out from the line of centers of the adjacent legs. The other leaf is horizontal, with its center 1 ft. out from the line of the legs on its side; and it has a 15-lb. pile of books resting in the middle of it. The table itself, minus leaves, weighs 30 lbs. Find the force between each leg and the floor.

19. An airplane in uniform horizontal motion is acted on by three forces. Its weight acts downward; the push of the air on its wing surfaces acts perpendicularly to them; and the force furnished by the propeller is horizontal. Draw these forces acting on a wing, assuming them to meet in a point, and explain how they balance, so as to leave no net force in any direction.

20. A simple truss-bridge has a span of 80 ft. and its weight averages 1 ton per linear ft. of span. A 3-ton truck stands 20 ft. from one end, and a 30-ton trolley car 30 ft. from the other. What are the forces acting on the two piers of the bridge?

21. If a wagon must be pushed along a road by hand, is it easier to move when one pushes on the body, or on the top of one of the wheels? Explain.

CHAPTER 4

MOTION

Uniform motion in a straight line. The motions of familiar objects, such as automobiles, footballs, house flies, etc. are really quite complicated because their speed is undergoing variations of a more or less irregular sort. We shall have to begin the study of moving bodies by considering a much simpler case, one which almost never happens, namely a small body moving with ***uniform motion in a straight line.***

An automobile moving uniformly along a level straight road has a ***speed*** given by the speedometer, and there is a simple relation between the distance s covered in a given time t and that speed. If we call the speed v, this relation is: the distance is equal to the product of the speed and the time, or $s = vt$.

A formal distinction is made between the terms ***speed*** and ***velocity.*** Speed is taken to be the numerical or ***scalar*** part of the velocity, i.e. the speedometer reading, without any thought of direction; whereas the word "velocity" is used for the ***vector*** quantity, in which both speed and direction are considered. This distinction is useful in certain cases, as examples will show. It is possible for an automobile in rounding a curve to keep its speed constant while its velocity is continually changing. Traffic officers are interested in the scalar quantity, speed, but travelers who wish to reach a given point are more concerned with the vector quantity, velocity.

Combinations of velocities. A body may possess two velocities at once, or even more. As examples consider an orange rolling across the floor of a moving train, or an aviator flying north in a

west wind. These velocities may be added geometrically in exactly the same way as has already been done with forces in statics, and since the process is identical in the two cases it should need no further explanation.

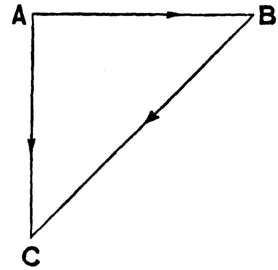


Fig. 4–1

Example of two velocities combined. A football moving horizontally with a velocity of 30 ft./sec. is turned through a right angle by a kick without altering its speed. Find the velocity which the kick alone would have given it.

We draw the velocity of 30 ft./sec. as a vector quantity AB, Fig. 4–1, on some convenient scale. AC must be the resultant velocity, after the kick, and it must be as long as AB but at right angles to it. BC is then the desired velocity which, when added to AB, according to the polygon method of combining vector quantities, produces AC. On this scale it is easy to see that its amount is $30 \times \sqrt{2}$, or 42.4 ft./sec., and its direction is at 45° to AC.

PROBLEMS

1. An automobile travels 20 mi. along a straight road at 25 mi. per hour. Another starts from the same place 15 min. later and maintains a speed of 35 mi. per hour. Which arrives first?

2. A man walks at a rate of 4 mi./hr. in a direction 30° N of E. How fast is he traveling northwards?

3. A power boat capable of making a speed of 6 mi. an hour in still water crosses a river which flows at the rate of 3 mi. an hour. (*a*) In what direction and how fast will the boat go if it is steered straight across the river by compass? (*b*) How should it be steered in order to go straight across, and how fast will it then travel?

4. A power boat is to cross a river on a course making an angle of 30° with the bank downstream. The speed of the river is 3 mi./hr., and of the boat is 12 mi./hr. At what angle must the boat be steered? (Solve by graphical method.)

5. Raindrops make tracks on the windows of a moving train which are at 30° to the vertical when the train has a speed of 10 ft./sec. What is the vertical speed of the drops?

6. A man walking in the rain holds an umbrella whose forward edge is 5 ft. from the ground. He walks at the rate of 5 ft./sec., and the rain is falling vertically at the rate of 20 ft./sec. How far ahead must the edge of the umbrella be if it is to keep his feet dry?

7. The smoke of a railroad engine blows due north when the train is at rest. When it is moving westward at a speed of 30 mi./hr., the smoke makes a

line which at a little distance from the train is actually northeast. How fast is the wind? In what direction do the smoke particles actually move? In what direction will the line of smoke seem to be as seen by the engineer?

8. An airplane is directed due north with a speed, in still air, of 60 mi. per hr. The wind is blowing from the northeast. The actual course of the plane is observed to be northwest. Find the velocity of the plane over the ground, and also that of the wind.

9. A man throws a package from a train moving with a speed of 12 ft. a second, so that it travels at right angles to the track at the rate of 9 ft. a second. With what velocity with reference to the train did he throw it, and in what direction?

10. A boy on a moving train hits a fence post with a bullet from an air rifle. At the moment the bullet leaves the rifle, the fence post is ahead at an angle of 45°. If the speed of the bullet alone is twice that of the train, show graphically how the rifle must have been aimed.

11. Two automobiles are approaching each other along a level road, each moving at the rate of 5 ft./sec. Before they pass, a passenger in one throws an orange out at right angles to his motion, and a stationary bystander observes that its path in the air makes an angle of 30° with the line of motion of the car from which it came. The orange is caught by a passenger in the other car. What was its speed with reference to him? (Solve graphically.)

Uniformly accelerated motion in a straight line. Bodies in motion usually change their velocities continually. These changes may be irregular and complicated, but there is a common type which is simple, namely that exhibited by ***falling bodies.*** If we neglect small effects due to friction, the characteristic feature of the motion of falling bodies is that their velocity is continually increasing, and that the rate of this increase is ***uniform.*** The rate of change of velocity is called ***acceleration*** and is denoted by a. It is identical with the "pick-up" of the automobile salesman. Acceleration is regarded as positive when the velocity is increasing. If the rate of change of velocity is constant, $a = v/t$, or $v = at$, where v is the velocity gained during the time t, starting from rest. This equation may be taken as the definition of a. Both v and a are vector quantities, and involve the idea of direction. The equation, of course, yields the speed (without reference to direction) if the direction is not specified.

Distance covered in uniformly accelerated motion. If a body moves according to this rule, the distance s which it covers in a given time, from rest, may easily be obtained. The initial velocity being zero and the final velocity at, the average velocity

has been $\frac{1}{2}at$ during this interval. If we substitute for the real body an imaginary one, moving with a ***uniform*** velocity of $\frac{1}{2}at$, this body will cover the same distance in this particular time that the real one does. But, by the law of uniform motion, this distance is the product of the velocity and the time, or

$$\begin{aligned} s &= \tfrac{1}{2}at \times t \\ &= \tfrac{1}{2}at^2. \end{aligned}$$

This means that in the first second (from rest) a uniformly accelerated body goes a distance equal to $\frac{1}{2}a$. In the first two seconds the distance is four times as great; in the first three, nine times as great, and so forth. In other words, the distances traversed vary as the squares of the times. From these the distances traversed in ***successive*** seconds may easily be found, if desired.

Equations of uniformly accelerated motion. A third equation is obtained if we substitute $t = v/a$ in the equation $s = \frac{1}{2}at^2$; which turns out to be $v^2 = 2as$. Likewise we may get a fourth equation by eliminating a from the first two. These four equations

$$\begin{aligned} v &= at \\ s &= \tfrac{1}{2}at^2 \\ v^2 &= 2as \\ s &= \tfrac{1}{2}vt \end{aligned}$$

are known as the ***laws of uniformly accelerated motion.*** For falling bodies they are often stated with g instead of a, as explained below.

If a body has a velocity v_0 at first, instead of being at rest, these equations take the forms,

$$\begin{aligned} v &= v_0 + at \\ s &= v_0t + \tfrac{1}{2}at^2 \\ v^2 &= v_0^2 + 2as \\ s &= \tfrac{1}{2}v_0t + \tfrac{1}{2}vt \end{aligned}$$

The derivation of these equations is left to the reader. They were worked out in essentially this form by Galileo, as a result of his experiments on inclined planes.

Acceleration of falling bodies. The acceleration of a falling body is usually called g (from gravity). Its amount is readily found by experiment, (p. 233). We shall use for the present the approximate values 32 (in feet and seconds) and 980 (in centimeters and seconds). More exact values are given later on, (see p. 113).

A little explanation is needed in regard to the ***units*** in which these numbers are expressed.

A velocity is always a distance divided by a time; e.g., feet per second, miles per hour, etc. An acceleration must involve a statement of time ***twice.*** Thus an automobile which acquires a speed of 30 miles an ***hour*** in 10 ***seconds*** could be said to have an acceleration of 30 miles per ***hour*** per 10 ***seconds***, or 3 miles per ***hour*** per ***second***, which works out to be 4.4 feet per ***second*** per ***second***, meaning that in each second a gain of speed of 4.4 ft./sec. occurs. A convenient abbreviation of this expression is $a = 4.4$ ft./sec.^{2} None of the units of velocity or acceleration have names. In such cases, which are rather frequent in physics, the custom is to state the units along with the numerical values in the manner just done.

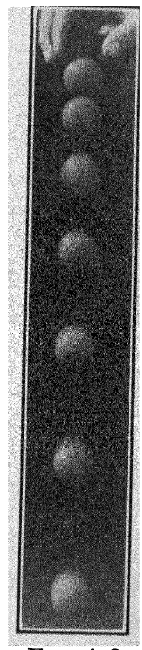


FIG. 4–2

A photographic record of a falling body, taken at regular intervals, about 22 per second.

If g is equal to 32 ft./sec.^{2}, this means that the velocity of a falling body at the end of the first second is 32 ft./sec.; but the distance traversed in that time is only 16 ft. In the first two seconds a distance of 64 ft. is covered and a velocity of 64 ft./sec. is acquired. If a body is thrown upward, it rises to a certain height, and then starts downward from a momentary state of rest. In the absence of friction, the downward journey is precisely like the upward one, but reversed. Thus, if a ball is thrown straight up with a velocity of 64 ft./sec., it will rise for 2 seconds, reach a height of 64 ft., and then turn and repeat the trip in reverse order, arriving at the bottom with the same speed as that with which it started. If it is ***any*** number of seconds in the air, it must rise for half that time and fall back to the starting point during the rest; in the rising half, the simple formulæ of "falling" bodies apply, if one remembers that the velocity is continually diminishing instead of increasing. One must be sure to break a problem in two if it involves both rising and falling, or else to use the more complete formulæ involving v_0 (p. 46) with care as to the proper sign given to the initial velocity.

In Fig. 4–2 is shown a trace of a falling golf ball photographed in sunlight through regularly spaced holes in a uniformly rotating

disc. Thus intermittent glimpses are obtained; the distances between the successive positions of the ball steadily increase and the length of each little image indicates the velocity at that position.

Example on falling bodies. A ball is thrown upward with a velocity of 80 ft./sec. Find the velocity and position of the ball after 4 sec.

(*a*) If we split the problem, we find first how long a time the ball was rising. From $v = at$, $80 = 32t$; or $t = 2.5$ sec. Hence it rises for 2.5 sec. and falls for 1.5 sec. It rises to a height given by $s = v^2/2a = 6400/64 = 100$ ft. It then falls for 1.5 sec. through a distance given by $s = \frac{1}{2}at^2 = \frac{1}{2} \times 32\ (1.5)^2 = 36$ ft.; and acquires a downward velocity of $v = at = 32 \times 1.5 = 48$ ft./sec. Its actual height above the starting level will be $100 - 36 = 64$ ft. at the end of 4 sec.

(*b*) Solving by the enlarged formulæ, the final velocity $v = -v_0 + at = -80 + 32 \times 4 = -80 + 128 = 48$ ft./sec. downward. Note that we take the direction of the acceleration as positive, which forces us to take the upward initial velocity as negative. The final distance from the starting level is $s = -v_0t + \frac{1}{2}at^2 = -80 \times 4 + \frac{1}{2} \times 32 \times 16 = -320 + 256 = -64$ ft., or 64 ft. upwards, since downwards was assumed as the positive direction.

Effects of friction on falling. If we take account of the friction of the air and make careful measurements on actual falling bodies, we find that they are very appreciably affected by it, even a ball made of material as heavy as lead showing a measurable deficiency in speed in falling from any considerable height. Lighter objects are so much held back that they cease to be accelerated at all, and fall with a constant speed after the first increase. A falling feather reaches a steady speed almost at once; a balloonist coming down with a parachute falls at constant speed after the parachute opens. Raindrops would be almost dangerous if they fell according to the simple "laws of falling bodies"; their actual final speed depends on their size, the minute ones in fine mist, or clouds, fall very slowly indeed. Ordinary drops travel at a rate which cannot be greater than 8 meters (26 feet) per second; a higher rate breaks them up into smaller drops. There is a definite law, known as Stokes' law, connecting the size of spherical particles and their rate of fall, which is often used as an indirect means of measuring the size of very small drops (p. 309).

PROBLEMS

1. A stone dropped from a bridge over a deep ravine strikes the bottom in 3.2 sec. as measured by a stop watch. How high is the bridge?

2. If a man throws a ball straight up in the air so that it returns in 4 sec., how high does it rise, and with what speed does it start?

3. How far does a body fall in $\frac{1}{4}$ sec., $\frac{1}{2}$ sec., $\frac{3}{4}$ sec.; how fast is it moving at these same times?

4. A boy throws balls up into the air one after the other, starting each one as the previous one reaches the top of its flight. If he can throw two a second, how high must each ball rise? How long a time does it take to rise and fall back to its starting level?

5. Two balls were dropped from different heights but reached the ground at the same instant. One took 1 sec. to fall; the other 2 sec. Where was the second ball when the first started?

6. A bullet from a small rifle travels at the rate of 300 meters a second. If there were no air friction, how far would a bullet have to fall in order to acquire this speed?

7. A rifle bullet starts with a speed of 400 m. a second and has a negative acceleration, due to air friction, of 15 m./sec.^{2} Assuming its motion to be entirely horizontal, how far will it go in 3 sec.?

8. A rifle bullet is shot vertically up into the air with a speed of 300 m. a second. It is retarded by an average (negative) acceleration of 10 m. / sec.^{2} due to friction in addition to the usual effect of gravity. How high will it rise in 10 sec.?

9. A local passenger train is just starting from a station when a freight train moving at a uniform speed of 40 ft. a second goes by on a parallel track. If the passenger train maintains a uniform acceleration of 2 ft./sec.^{2}, after what time will the trains again be opposite each other? How fast is the passenger train then moving, and how far is it from the station?

10. How long would it take an automobile to move without friction down a hill 100 ft. high, inclined at 30° to the horizontal?

11. A boy stands on a bridge, 36 ft. above the water. He drops a stone. With what speed does it hit the water? He then throws another stone down, starting it with a downward speed of 8 ft./sec. How long is it in the air, and what is its speed when it hits the water?

12. A carpenter stands on an elevated platform, 36 ft. above the street level. He throws a hammer up to an assistant 12 ft. above him who misses it, so that it falls to the street below. How long was the hammer in the air?

13. A sled slides down a smooth hill which falls 2 ft. vertically for each 10 ft. along the slope. It reaches the bottom in 10 sec. Find the acceleration and the length of the slope.

14. A sled slides down a smooth hill 100 ft. long, and reaches the bottom in 10 sec. Find the acceleration, the position of the sled in the middle of the time, and the time when the sled was halfway down.

15. A balloon is ascending with a uniform velocity of 16 ft./sec. A bag of sand is released (not thrown either down or up) at a height of 140 ft. from the ground. Trace the subsequent motion of the sand, finding times and distances.

PROJECTILES

Path of a projectile. A body thrown horizontally continues to move horizontally, while at the same time it falls like a falling body, becoming what we call a "projectile." It is curious, but true, that neither of these motions affects the other in the least. Supposing no friction to interfere, the horizontal motion remains constant at its initial speed, v_0, while the body is in flight. The body will at any time t be found at a horizontal distance x away from the starting point which is given by the law of uniform speed, $x = v_0t$. At the same time it will have fallen a vertical distance y downward which is identical with that which it would

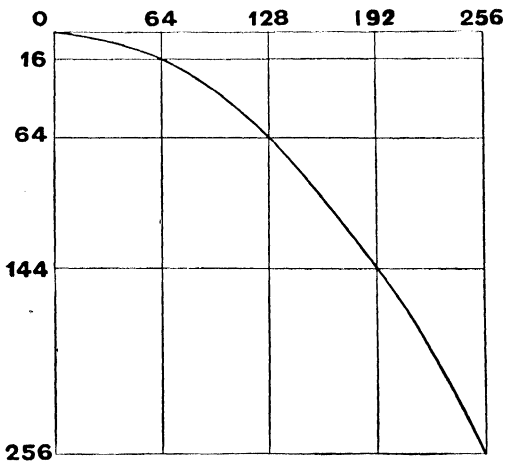


FIG. 4–3

have covered if it had not moved sideways at all. This distance is given by $y = \frac{1}{2}gt^2$. Since the time t is the same in both equations, we may eliminate it, and if we do we get the equation $x^2 = 2v_0{}^2y/g$. This means that the horizontal distance x and the vertical distance y are related so that x^2 is proportional to y. It follows that the body traces a curve which is known as a parabola. Its shape is familiar to anyone who has thrown a ball.

A simple way of seeing how it goes is to draw a series of equidistant vertical lines as in Fig. 4–3 separated by a distance of v_0, (assumed to be 64 feet per second in the case shown in the figure) the first one passing through the starting point of the projectile. The body will then be found to be on each of these lines successively at intervals of one second. Horizontal lines may then be drawn

below the level of the starting point at distances corresponding to 16, 64, 144 ft., etc., these being the distances reached by a falling body in successive seconds. Where each horizontal line crosses the corresponding vertical line the body will actually be found at the time to which the lines refer.

An interesting way of tracing such curves is by rolling a heavy round ball over a piece of paper resting on a slightly inclined glass plate. By first covering the white paper with thin carbon paper, a permanent trace is left which will be approximately parabolic

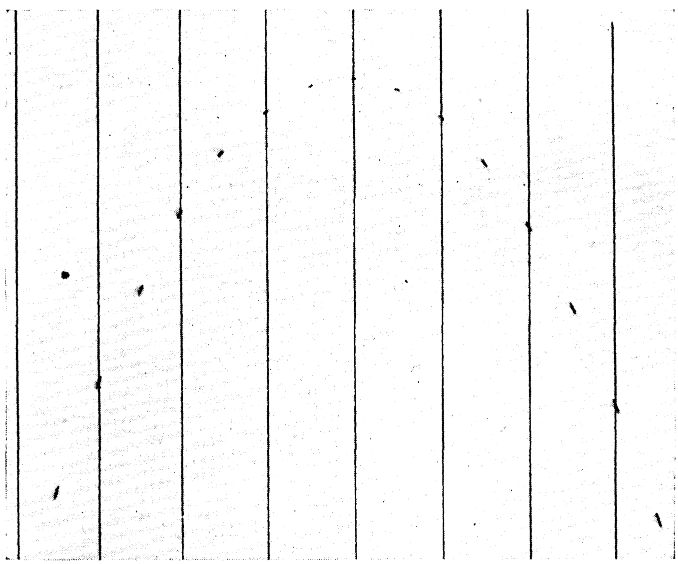


FIG. 4–4
Photographic trace of the path of a projectile

in shape because the ball is "falling" on the inclined plane, though at a slow rate. Another method is to photograph a projectile while it is in flight, using a camera which sees it through a rapidly revolving disc containing a few regularly spaced holes. In this way Fig. 4–4 was obtained, on which uniformly spaced vertical lines have been drawn which show plainly the constancy of the horizontal motion.

Heights and ranges. If a body is thrown upward with a velocity v at an angle θ with the horizontal, it will have a uniform horizontal velocity of $v \cos \theta$, and a varying vertical velocity starting at the value $v \sin \theta$. The height to which it will rise will be obtained from the equation $(v \sin \theta)^2 = 2gh$; and

the time spent in rising will be given by $v \sin \theta = gt$. The whole time of flight is twice this amount. The range, i.e., the total horizontal distance covered, is the whole time of flight multiplied by the horizontal velocity, $v \cos \theta$. These conclusions hold only if friction is negligible. In the case of real bullets, long-range shells, etc., they fail seriously. Friction makes the path quite different, and reduces the range to less than half what it would be in a vacuum.

Position of projectile by vector addition. An interesting way of finding the position of a projectile is by combining its two displacements as vectors. Thus if a stone is thrown as in Fig. 4–5 in a direction AB (which need not be horizontal) with a velocity v_0, it would have passed over in time t a distance v_0t along AB, if gravity had not acted on it. If it had merely been dropped from A, it would have fallen a distance $\frac{1}{2}gt^2$, which is AC in the figure. As a result of these two displacements it must actually be at the point D at the end of time t, having followed the curved path shown in the figure.

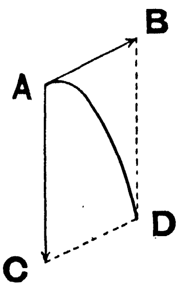


FIG. 4–5

Deflection of projectiles by rotation of the earth. The rotation of the earth produces a curious effect on projectiles which is detectable on long ranges. If a rifle is fired due northward from a point in the northern hemisphere, the bullet has, while it is in the rifle, the same eastward velocity that everything else has at that place, due to the fact that the earth is rotating so as to make its surface move eastward. The bullet does not lose this eastward velocity merely because it is discharged. Hence it goes northward because of the action of the rifle and eastward because of the velocity of the earth at its point of departure. But the target, being farther north, is not moving eastward at quite so fast a rate. Therefore the bullet deviates slightly to the right of the place aimed at, by an amount which increases with the range and with the duration of flight. Similar reasoning shows that the deviation is to the right, whatever the direction in which the bullet is fired; but in the southern hemisphere the deviation is reversed, and at the equator it does not occur at all. At a place in 45° north latitude a rifle bullet fired north over a range of 1000 yards would be deflected nearly 4 inches if it took 2 seconds to make the journey, and more in proportion if it were slower. Precisely the same cause is responsible for the fact that air starting to flow northward is deflected toward the east. Thus the air which de-

scends from high levels at about latitude 30°, in the northern hemisphere, and starts to flow northward and southward from there is deflected to the right in both cases, becoming a southwest current over the region to the north and a northeast ("trade") wind over the region to the south. The general drift of the air in a northeasterly direction north of 30° is responsible for the travels of the storm areas already mentioned (p. 18), and also for the unusual difficulties experienced by those who try to fly across the Atlantic Ocean from Europe to North America, compared with those met with by aviators going in the opposite direction.

PROBLEMS

1. A football was timed on a long kick and observed to be in the air 5 sec. How high did it rise?

2. A small rifle is mounted on a platform 24 ft. above the surface of a lake and its barrel is carefully leveled. The bullet when fired strikes the water 400 ft. away from the base of the platform. Find the velocity of the bullet as it leaves the rifle. (Neglect air friction).

3. A stone is thrown horizontally from a bridge with a velocity of 10 ft./sec. and strikes the water below in 3 sec. How high is the bridge? How far away horizontally did the stone strike? How fast and in what direction was the stone actually moving when it struck? (Specify the direction by a diagram.)

4. A weight is allowed to fall 4 ft. in a moving train. Find the time required for this fall in case the train is (*a*) moving uniformly at a speed of 10 ft./sec.; (*b*) moving with an acceleration of 2 ft./sec.^{2} Show by a sketch the path of the falling weight as seen in each case by an observer in the train.

5. An aviator drops a heavy object from a height of 1600 ft., while he is moving uniformly in a horizontal line at the rate of 90 ft./sec. How long is it before it strikes the ground? Where is the airplane with reference to the object when it strikes? How far does the object strike ahead of the point which was directly under the airplane at the moment when the object was released? (Neglect air friction.)

6. A man throws a ball horizontally from the roof of a building 80 ft. high so that it strikes the ground across the street, a distance of 100 ft. away horizontally. With what speed did he throw the ball?

7. An automobile traveling at 30 mi. an hour along a level road crosses a 2-ft. ditch in the road, which has been filled up to within half an inch of the level surface of the road. Does the car strike the filled-in surface, or does it "jump" from one edge to the far side without touching? (Assume that the

wheels act as freely falling bodies and that the tires are not flattened at the point of contact with the road.)

8. A wild duck rising from the water patters along the surface and raises little jets of water with each foot. A quick photograph shows that the water in these jets rises to a maximum height of 27 cm. Assume the jets to be all alike. The picture (Fig. 4–6) shows the duck to have reached a point 120 cm. beyond the particular jet in which the water is caught by the camera at its greatest height. Find the average speed of the duck's flight in this interval.

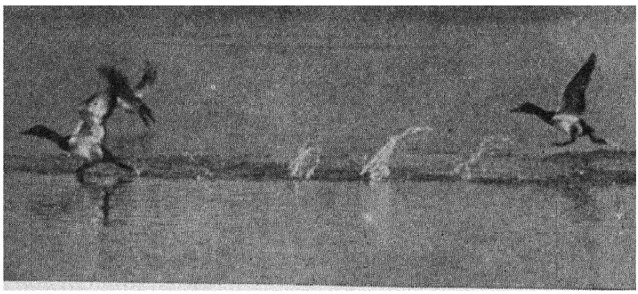


FIG. 4–6
Photograph by Professor A. A. Allen of Cornell University

9. What is meant by a uniform acceleration? Is the acceleration of gravity correctly stated as 980 cm. per second? If a heavy object falls from the ceiling of a railway car while the car is being uniformly accelerated, what sort of a path (i.e., curved or straight, forward or back) does it follow as seen (*a*) by an observer in the car, (*b*) by a stationary observer outside?

10. From the equations of p. 51 derive the range of a projectile which starts at an angle of 45°.

11. Work out the flight of a shell whose range is about 60 miles, say 320,000 ft., assuming no air resistance. Assume it to start at a 45° angle, and find the velocity, the time of flight, and the height to which it rises. (When a real shell is sent over as long a distance, the air resistance makes a great difference, but not so much as one might expect, since it is very small in the upper part of the flight.)

UNIFORM MOTION IN A CIRCLE

Acceleration in uniform motion in a circle. The case of a small body revolving uniformly in a circle introduces us to some curious effects of considerable importance in all cases of rotation. Is uniform motion uniform in every sense or not? Here we find a good use for the idea of a vector quantity, and for the distinction between speed and velocity. An ordinary uniform velocity in a straight line is a true vector quantity, and is completely repre-

sented by a straight line of a definite length. Evidently, if we try to represent uniform motion in a circle by any one line we shall be in doubt as to how to draw it for the reason that the direction of the motion is continually changing. The scalar part of the velocity, which we call the speed, remains constant, but the velocity changes since the direction changes. Is the body accelerated, then, if the velocity is changing? Certainly it must be, or our definitions have been inconsistent. Acceleration is then possible even when speed remains constant, for a change in direction is just as truly a change as is a change in amount.

Calculation of the acceleration. We shall proceed to find the acceleration of a small body revolving uniformly in a circular path. If the body is at P (Fig. 4–7) at one instant it will reach

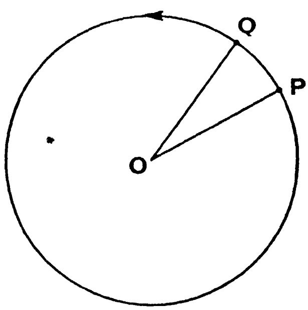


FIG. 4–7

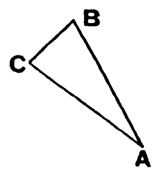


FIG. 4–8

the point Q in a very short time t, afterwards. The line PQ is supposed to be an extremely short distance, drawn much too large in the figure; being so small, the arc PQ is not appreciably curved.

At the instant the body is at P its velocity is along the tangent to the circle at that point. Let AB (Fig. 4–8) be a vector drawn to represent the velocity at P; i.e., the length AB represents the speed and the direction of AB is parallel to the direction of motion of the body at P. When the body reaches Q the speed is the same but the direction has become parallel to AC; the line AC then represents the new velocity. Consider what must have been done to change the velocity AB into AC. Evidently, if we add (geometrically) the velocity BC to AB we shall get AC; or, we may say BC is the (geometrical) difference between AB and AC. If a body is changing its velocity it has an acceleration a, and in the time t (here very short) the change of velocity is equal to at. This must then be what BC represents; i.e., $BC = at$.

Consider next the triangles QOP and CAB. The narrow angles are equal and each triangle has two equal sides. Therefore they are similar, and we may say that $PQ/OP = BC/AB$. But PQ is the distance covered with velocity v in a time t; or $PQ = vt$. OP is the radius r of the circle; AB is drawn equal to v, and BC is equal to at. Hence, $vt/r = at/v$, or

$$a = \frac{v^2}{r}.$$

This gives us the amount of the acceleration in this type of motion, which we shall find useful later. The ***direction*** of the acceleration is evidently along the radius of the circle, for it is parallel to BC, which is perpendicular to the tangents AB or AC; (it does not matter which we say, since the angle between them is vanishingly small).

PROBLEMS

1. A ball is rolling around in a circular groove of 20 cm. radius in a horizontal table with a uniform speed of 7 cm./sec. Find its acceleration.

2. A passenger riding in a closed automobile is rounding a curve whose radius is 40 m. at a speed of 4 m. a sec. If the road is level, how would a lead ball seem to fall inside this automobile when watched by the passenger? Would its path be straight or curved, and in what direction would it appear to go? How would it appear if watched by a stationary observer outside the car? How would the ball hang if suspended as a pendulum from the top of the car?

3. A vertical rod is made to rotate uniformly about its own axis. A small ball is hung on a string attached to the top of the rod, and stands out so that the string makes an angle of 45° with the rod. If the string is one meter long, what is the speed of the ball?

4. In the last problem, find the algebraic connection between the speed of the ball and the length of the string (assuming an angle of 45° always).

5. A vertical post is rotating once in 2 sec. To the top of it is fastened a ball by means of a cord 6 ft. long. After a steady state is reached the ball stands out from the post a certain distance and revolves with it. Find this distance.

CHAPTER 5

FORCE AND MOTION

The laws of force. Galileo was the first to adopt the experimental method and thus to make direct inquiries from Nature itself. His predecessors, following the Greek philosophers, thought the human mind ought to be powerful enough to reason out all natural phenomena, and regarded it as somewhat beneath their dignity, or perhaps unsportsmanlike, to make experiments. Galileo presents in his "Dialogues," (1632) [1] many sound ideas about the mechanics of moving bodies. He first saw clearly the constancy of the acceleration of falling bodies, the importance of momentum (p. 65) and its relation to force. The distinctions among force, work, energy, etc., were slow in coming and in being universally agreed upon. It was not until the nineteenth century that our present science of mechanics was firmly established, but a great advance is to be credited to Newton.[2] The foundation which he

[1] "Dialogues Concerning Two New Sciences" by Galileo Galilei; translated by H. Crew and A. de Salvio (Macmillan), 1914; a book giving an interesting glimpse of the state of science in those days.

[2] Sir Isaac Newton (1642–1727), English experimental and mathematical physicist, whose scientific successes gave him a position of extraordinary prestige and influence, so much so in fact that in the few cases in which he was wrong it took a century or more to set matters right. Perhaps his greatest achievement was in formulating and testing his law of gravitation, though his "Principia," published in 1687, greatly increased man's grasp of mechanical principles in general. His work on optics was very extensive and included the discovery of the spectrum of white light, a careful study of the colors of

laid for mechanics is briefly put in the form of three laws, a free statement of which follows:

1. ***A body remains at rest, or in a state of uniform motion in a straight line, unless a force acts on it.***
2. ***Force creates momentum in its own direction, and is measured by the rate of change of the momentum created by it.***
3. ***To every force there is an equal and opposite reaction.***

These are known as "Newton's laws of motion," though he states in his "Principia" that the first two of them were given by Galileo. The meaning and consequences of these three laws will now be taken up in turn.

Newton's first law. This law has many aspects. First, it implies that it is natural for all bodies to remain at rest. In earlier times this was thought to be the only natural state of bodies, since ordinary motions do not actually last very long; but the reason for this is, of course, that friction is present, opposing the motion, and friction is itself a force. If no force whatever acts on a body, after it has already been set in motion, it will continue to move in a straight line indefinitely, and the law implies that a state of uniform motion in a straight line is just as natural and reasonable as a state of rest, even though it almost never occurs as a reality.

The law also indicates that force is the sort of action whose essential character is its ability to make bodies move, or to change their motion. In other words, a force produces a change of velocity, or an acceleration. This we may regard as a qualitative definition of what we mean by force, though as yet we have said nothing about exact means of measuring it.

Forces in real life appear as weights, or are caused by muscular exertion, the distortion of elastic bodies such as stretched springs, the pressure of steam in boilers, or by a variety of other conditions. They produce some effects, such as muscular fatigue, which

---

thin films, and of the phenomena of double refraction, polarization, etc. He made excellent reflecting telescopes, and was much interested in lenses. As Master of the Mint he was led to study alloys and high temperature methods. He also wrote several religious works. He was Professor of Mathematics at the University of Cambridge, and invented a method of calculation essentially equivalent to the differential calculus, which involved him in a famous controversy over priority with Leibnitz.

are not measurable, and others, like changes in the motions of bodies, which are; the latter are among the simplest force effects to observe, and hence the best from which to start in devising methods for the exact definition and measurement of force.

Inertia. Another aspect of Newton's first law is the important part played by the property of inertia, which is common to all matter. We all know from common experience how hard it is to start a heavy body moving, or to stop it once it is in motion. This reluctance to change is due to its inertia. Several experiments may be mentioned to show the nature and effects of inertia.

If one makes a pile of half a dozen similar wooden blocks, one may bring another block along the table and by a smart blow with it against the bottom block of the pile knock the latter away, and substitute the new one in its place without greatly disturbing the other blocks. Their inertia keeps them at rest, or nearly so, because the feeble force of friction between the blocks is called into being only during the short time of the collision; and in this time it cannot move the blocks much on account of their inertia. The large force on the bottom block, however, is able to give it a high velocity in the same time. If we bring the free block gently against the pile, we can push it as a whole about the table, there being time now for the force of friction to give all the blocks the same speed.

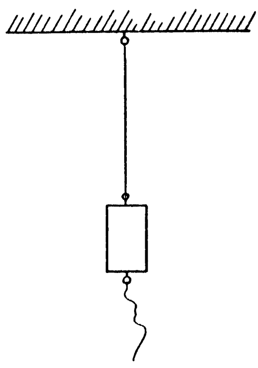


FIG. 5–1

If one hangs a rather heavy weight on a thread, with an extra piece of thread below, (as in Fig. 5–1) a steady downward pull on the lower thread will break the upper one first, because it must sustain the heavy weight in addition to the pull; while a quick downward snap will usually break the lower thread only, the upper one being momentarily protected from much stretching by the inertia of the weight.

Measurement of inertia. Such experiments as those above illustrate the nature of inertia. If we wish to measure it, however, we need a unit of measurement, and a quantitative experiment with which to test it. Let us imagine, for example, a level, frictionless surface, a body resting on it, and an extraordinarily good and sensitive spring balance (p. 125) with which we can somehow maintain a steady pull on the body along the surface, exerting

always exactly the same force. We could then, at least in imagination, measure the acceleration of the body, say by counting the time it takes to pass over a measured distance. Now, keeping always this same force, we might apply it to different bodies. Those with the same accelerations are said to have the same inertias. Those that move with half as much acceleration have twice as much inertia, and so on. In other words, in such an experiment, the inertia varies inversely as the acceleration. Thus we reach a quantitative definition of inertia which corresponds with the accepted meaning of the word, and is really included in Newton's second law. Inertia is a property of a body which it would presumably retain unchanged if the body were taken off into empty space, entirely away from the action of the earth. Under these conditions a body would have ***no weight,*** for we define the weight of a body as the force with which the earth pulls upon it; but we should expect it to be just exactly as hard to start or to stop as before.

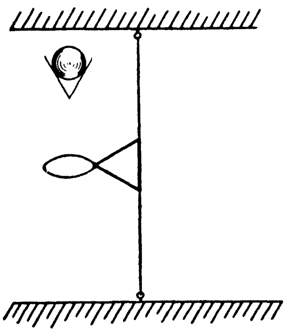


Fig. 5–2

The inertia balance. The apparatus called the ***inertia balance*** enables us to compare inertias ***without regard to weights,*** as in the last paragraph, but with more accuracy. A tightly stretched vertical steel wire supports a stiff, light frame holding a ring, placed a little distance (say 5 cm.) to one side of the wire, as shown in Fig. 5–2. A light cone, perhaps of stiff paper (shown separate in the figure), may be set with its apex down into the ring. A ball dropped into this cone will always be at a fixed distance from the wire, whatever its size. If, then, the ring so loaded is pushed aside and let go, it will oscillate around the wire, moving always in a horizontal plane, and therefore quite unaffected by any vertical forces, such as weights. We can see without going into details that the stiffness of the wire controls the motion, and that the more inertia the moving parts have the slower will be the rate of the oscillation. We must not infer that the one is proportional to the other, but there is a definite relation between them. Hence, by timing the motion, we can compare the inertia of different balls, and this can be done with some accuracy.

Mass and its units. Inertia, as illustrated by such experiments, is one of two fundamental attributes of matter which are together implied under the name of ***mass.*** The second is the ability that masses have to attract one another by gravitation (see Chapter 8) of which the weight of a body on the surface of the earth is a special case. The mass of a body cannot be more definitely described than to say that it is that quality to which are due both its inertia and its gravitational action. The mass is sometimes described as the quantity of matter in the body. This description is useful so long as we compare quantities of the same sort of matter, but fails to have a definite meaning for other cases.

Mass is measured by inertia experiments, or by weighing, in terms of arbitrary units, called standard masses, such as the International Standard Kilogram, preserved with elaborate precautions at the International Bureau of Weights and Measures at Sèvres, near Paris. The gram is one-thousandth part of this; it was intended to be, and almost exactly is, equal to the mass of a cubic centimeter of water at the temperature at which it is most dense (see p. 155). The "standard pound" is similarly preserved in London and serves as a standard for Great Britain, but in the United States the pound is legally defined in terms of the kilogram. Multiples and submultiples of these units make "sets of masses," which we commonly call "sets of weights," since mass and weight are closely connected.

Mass and weight. It is in consequence of a body's mass that it possesses inertia and that the earth is able to pull upon it with a gravitational force which we call its weight.[1] The more mass it has, the more weight it possesses. One is proportional to the other. The most exact proof of this comes from experiments on the rate of oscillation of pendulums (p. 233). Weight is not an attribute of matter in general, but is one due solely to its presence on the earth's surface. Matter at the center of the earth has mass, but no weight, for the pull of the earth upon it then is equal in all directions. Similarly the earth as a whole has mass but no weight; and a body moving up away from the earth's surface has a diminishing weight which would vanish altogether at a very great distance. The inertia of a piece of matter, on the

[1] Strictly speaking, this statement is true only at the poles of the earth. Elsewhere the weight of any body is the gravitational pull of the earth upon it less the slight effect due to centrifugal force (p. 71).

(*b*) Force is proportional to rate of change of "momentum," (mv);

$$F \propto \frac{mv}{t}$$

which are identical, since $a = v/t$ and m remains constant. The statement (*a*) includes the above results about inertia, for it implies that if F is kept the same, ma is constant, or a varies inversely as m; while, in the second sort of experiment, where m is kept constant, a varies directly as F. The form (*b*) will be considered later on (p. 65).

The ratio form of the second law. The first form of the second law states that F is ***proportional*** to ***ma,*** but if the common units are used F is ***not equal*** to ***ma***. From this form of the law, however, it is easy to deduce an equation instead of a mere proportion.

A force F acting on a particular body of mass m gives that body an acceleration a. The special force which we call the weight of the body, W, gives this same body an acceleration, g, which is that possessed by all falling bodies. Since the force is proportional to the acceleration, this statement may be written as a ratio:

$$\frac{F}{a} = \frac{W}{g} \quad \text{or} \quad \frac{F}{W} = \frac{a}{g}$$

In the last form, which we shall call for convenience, the "ratio equation," F and W on the left side are alike in nature, each being a force, and they must be expressed in the same units. But these units may be anything we please, since a ratio of two like quantities has the same value whatever units are used to measure them. The mass has disappeared, and with it any possible confusion between mass and weight. On the right-hand side a and g must be in the same units, but not necessarily in the same system as that in which F and W are expressed. There is thus less chance of mistake in using this equation than in other forms. On this account it is strongly recommended for use in the solution of all problems involving the second law. Examples will illustrate its use and its advantages.

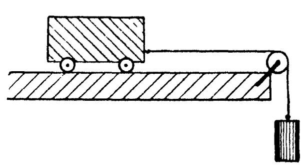


FIG. 5–3

Example on the second law. A weight of 100 gr. (Fig. 5–3) is resting on a level, frictionless table, and is tied by a light string to a 50-gr. weight hanging over the edge. Find the acceleration, and the tension (force) in the string. We note first that the force which produces the motion

is the 50 gr.; the weight of the 100 gr. can do nothing; it acts perpendicularly to the table, and the only possible motion is along the table. If it also were active (i.e., if both weights hung over the edge), the acceleration would, of course, be g. We have therefore by the ratio equation

$$\begin{aligned} F/W &= 50/150 \\ a/g &= a/980 \\ a/980 &= 50/150 = 1/3 \\ a &= 326.7 \text{ cm./sec.}^2 \end{aligned}$$

To get the tension, "isolate" the 100-gr. weight. It is being pulled by the tension T at an acceleration one-third that of gravity; so that, $T/W = a/g = \frac{1}{3}$, or $T = W/3 = 100/3 = 33.3$ gr. One might inquire what has become of the rest of the 50-gr. pull; evidently it is "used up" in producing the acceleration of the 50-gr. weight.

Another example. A 1000-gr. weight is hung on a spring balance in an elevator. Find the readings of the balance under different circumstances.

(a) Constant velocity of the elevator, either up or down, and of any amount. There is no acceleration, and therefore no force is required to maintain the motion. Of course, in reality a force was required to *start* the elevator, but that has ceased to act by the time the velocity has become constant; also, a force is required to overcome friction, but this is all used up in the process, leaving no *net* force. Hence there is no force F as contemplated in the ratio equation, that is, no force required to produce an acceleration, and thus the spring balance has nothing to do more than it has when the elevator is at rest. It reads 1000 gr.

(b) Uniform acceleration of 98 cm./sec.^{2} upward. The balance must now not only support the weight, but also give it an acceleration. The latter is what the force F does. This force is given by

$$F/1000 = 98/980, \quad \text{or} \quad F = 100 \text{ gr.}$$

The reading of the balance is thus 100 gr. more than usual, or 1100 gr. in all.

Absolute units of force. The dyne. There are other ways of regarding Newton's second law which lead to important consequences. Suppose we say that instead of making F ***proportional*** to ma, we prefer to choose such new units of F or of m as to make F ***equal to*** ma. If we imagine a unit mass (1 gm. or 1 pd.) being pulled along a smooth horizontal surface, and given a unit of acceleration (1 cm./sec.^{2} or 1 ft./sec.^{2}), then both m and a are unity; the equation is satisfied in this case if we call the force required to do this our new unit of force. Such a force is known as an absolute unit, and in the "C. G. S." (centimeter, gram, second) system it is known as the ***dyne.*** Thus the dyne is the force required to give a mass of 1 gm. an acceleration of 1 cm./sec.^{2} It is called an absolute unit because its definition would hold

equally well upon the surface of any other planet than the earth, or even in empty space, if physics experiments could be performed there with the centimeter, gram and second as units; that is, the definition is independent of local circumstances, such, for instance, as the variable pull of the earth. The dyne is related to the gram of force in a simple manner. Since 1 gr. acting on 1 gm. produces an acceleration of 980 cm./sec.^{2}, while 1 dyne produces an acceleration of only 1 cm./sec.^{2}, evidently 1 gr. = 980 dynes. This number varies slightly over the surface of the earth, as the gram-force is not constant. The dyne has the same value everywhere.

English absolute unit. The unit of the English system corresponding to the dyne is the ***poundal,*** the force required to give a mass of 1 pd. an acceleration of 1 ft./sec.^{2} It is not much used. Engineers usually dislike it because it is not "practical"; pure scientists do not use it because they prefer to use the more convenient C. G. S. system throughout; but it has enough importance to be worth mentioning. Reasoning as we did with the dyne, we see that (approximately) 32 poundals are equal to 1 pound-force.

Alternative unit of mass. There is still another way in which the equation $F = ma$ may be satisfied; we may retain the familiar gravitational units for force (lb. and gr.) and adopt new units for mass. This has never seemed worth doing in the metric system; in the English system, the ***slug*** is defined as a mass of 32 (nearly) pds. Then F lbs. = m slugs $\times$ a ft./sec.^{2} The slug is known as a "British Engineering Unit" of mass. So far as we know, British engineers avoid it, and, as we do not need it, we shall follow their example.[1]

Momentum. Collisions. There is still another aspect of Newton's second law that is worth considering. The form (*b*) on page 63, $F = mv/t$, is obtained from $F = ma$ by using the fact that $a = v/t$. The quantity mv, mass multiplied by velocity, is called the ***momentum*** of the body. It used to be called, sometimes, the "quantity of motion." This term is not used nowadays, but it is descriptive. If one considers the danger associated with a collision, it is plain that the momentum is what counts; a moving

[1] The ratio equation may be written $F = W/g \times a$. This is equivalent to putting W/g for m in the equation $F = ma$. This form of the second law seems to imply measuring mass in slugs (in the English system).

feather is never dangerous, while even a slowly moving automobile can exert an enormous force against an object that suddenly stops it, or checks its motion. The *forces* do the damage, and they are in proportion to the rate at which the momentum is being destroyed. Certain problems are conveniently solved by the use of the momentum form of the second law, though they can always be solved otherwise, if preferred. It is to be remembered that when we write $F = mv/t$, we mean by mv the ***change of momentum in the time t.*** If the body starts from rest, v is the velocity acquired; if not, v is the change of velocity produced by the force. In using this equation we should agree to use ***absolute*** units for F (dynes or poundals). The next example shows how easily this may be done. These units are not hard to use, if we have to; we are quite accustomed to changing pounds to ounces, or vice versa; here we have to change pounds to poundals (1 lb. = 32 poundals) or grams to dynes (1 gr. = 980 dynes); the process is no more difficult.

Example on collisions. A 2-pd. hammer, moving with a velocity of 10 ft./sec., hits a nail and drives it $\frac{1}{4}$ in. into a board. Find the average force between nail and hammer during their time of contact. The time of contact is given by the equation $s = \frac{1}{2}vt$ (p. 46), assuming that the motion is uniformly accelerated (negatively). $\frac{1}{4}$ in. $= 1/48$ ft. $= \frac{1}{2}vt = \frac{1}{2} \times 10t$, whence, $t = 1/240$ sec. The force F (poundals) $= mv/t = \dfrac{2 \times 10}{1/240} = 4800$ poundals, or $4800/32 = 150$ lbs.

If we prefer to solve this example by the use of the ratio equation, we can find the (negative) acceleration of the hammer by $v^2 = 2as$, or $a = \dfrac{10^2}{2 \times 1/48}$ $= 2400$ ft./sec.^{2}, whence

$$F/W = F \text{ (lbs.)}/2 = a/g = 2400/32,$$

from which

$$F = 4800/32 = 150 \text{ lbs.},$$

as before. This solution is a little less direct, but avoids mentioning absolute units of force.

If it had been stated in the example that the nail and hammer remained in contact through a time of 1/240 sec., the quickest solution would have been by

$$F = mv/t = 4800 \text{ poundals} = 150 \text{ lbs.}$$

Newton's third law. Newton's third law states that "to every force there is an equal and opposite reaction." This implies that when one body acts on another the second acts equally on the

first; or that within one body, if one part pulls on the rest of it, there is an equal and opposite pull in return. The truth of this statement is usually obvious. A man's weight is opposed by the upward push of the chair in which he sits; these two forces are equal and opposite. A weight hanging on a rope pulls down on the rope and is pulled up by it, equally; the rope pulls down on its support, while the support pulls up on it. Two well-balanced parties in a tug-of-war produce a state of tension in the rope which has obviously the same two-sided nature. Like buying and selling the two aspects of the one affair occur together; one cannot exist alone.

One may test the truth of this law by means of two spring balances which are graduated alike, as shown by testing them with equal weights. One cannot pull against the other without creating an equal reading in it. Other simple tests can readily be devised.

These are cases in which no motion occurs as a result of the action of the forces. The law is equally true though not quite so obvious in cases where accelerations occur. Suppose, for example, that a man pulls a cart after him by means of a rope; assume frictionless motion and a steady pull, starting from rest. The cart will then be accelerated uniformly, and the reader might say that we have here an exception to the rule because the cart yields, and gives way to the action of the force. It cannot on the frictionless surface "dig its heels in" and resist the pull. But, even in this case, the inertia of the cart, making it continually lag behind, produces a retarding force acting from the cart on the rope, ***exactly equal*** to the forward one. It is not possible for these two to be unequal; in a light cord the tension must always be the same in both directions and in all parts. If this seems odd, the reader may try, if he likes, the simple experiment of pulling with a large force, say 50 lbs., on one end of a thread which is lying ***loosely*** on a table. It cannot be done; the thread follows and prevents the force from ever rising to a large value. One could not pull strongly on the thread unless there were a weight on the other end capable of holding back on account of its inertia, and pulling on the thread with just the same force. Forces of the kind here referred to are called ***inertia reactions.*** If one pulls harder, the inertia reaction becomes greater because the acceleration is greater, and always action and reaction are equal and opposite. The inertia reaction exists only while the body is being accelerated. If the accelerating

force ceases to act (e.g. if the string breaks) the inertia reaction instantly vanishes.

An interesting example of the third law is furnished by a rifle bullet imbedding itself in a plank. The bullet cannot push on the plank any harder than the plank does on it. The bullet by pushing on the plank makes a hole in it; the plank by pushing on the bullet brings it to rest.

PROBLEMS

1. Could a rocket propel itself in a vacuum?

2. A tiger and a flea jump, each as far as it can. Which is exerting the larger force, the tiger in mid-leap, or the flea on landing?

3. Leonardo da Vinci stated the following propositions. Are they correct?

(*a*) If a force moves a body for a given time over a given distance, the same force will move half the mass in the same time through twice the distance,

(*b*) or the same force will move half this mass through the same distance in half this time,

(*c*) or this force will move twice this mass through the same distance in twice the time,

(*d*) or half this force will move half the body through the same distance in the same time,

(*e*) and half the force will move the whole of the mass through half the distance in the same time.

4. Could one set up a powerful blower in a sailboat and propel the boat by blowing on its sails? In what direction would such a device be likely to move the boat? Explain.

5. Descartes stated the following laws of motion (among others). Are they correct?

(*a*) All bodies strive with all their might to stay as they are.

(*b*) The measure of a body's force is the product of its mass and its speed.

(*c*) A moving body tends to keep the same speed and direction.

6. Aristotle's argument about falling bodies runs something like this: a brick falls at a certain rate; if another brick is laid upon it, the upper one pushes upon the lower one; therefore two bricks fall faster than one. Discuss.

7. A sled at the bottom of a hill has a speed of 20 ft./sec. It runs over a level surface for 20 sec. before stopping. Find the acceleration (negative) and the distance covered in this time. If the mass is 200 pds., find the force due to friction which brought the sled to rest.

8. Find the force with which a 150-lb. man pushes against the back of his seat in a car which has a forward acceleration of 3 ft./sec.^{2}

9. A weightless pulley supports two 200-gram weights (Fig. 5–4) as shown, by a light cord. An extra 50-gram weight is attached to one side. Neglecting friction, find the acceleration of each weight and the tension of the cord. (This instrument is called the Atwood machine, and is sometimes used to test the truth of Newton's second law. Friction and the inertia of the pulley usually introduce serious errors in the measurements.)

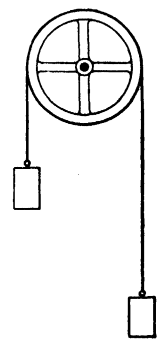


Fig. 5–4

10. In a modification of the last problem two monkeys of equal weight are substituted for the weights in Fig. 5–4. One is inactive; the other climbs up his cord toward the pulley. What happens to the inactive one at the same time? If the cord were heavy what difference would that make?

11. A locomotive weighing 300 tons is attached to a train of three cars, each of which weighs 100 tons. Find the horizontal force in tons that the engine must exert against the rails (through friction) if the train is to start with an acceleration of 1 ft./sec.^{2}; and also the force in the coupling by which the last car is attached to the train.

12. A man in a rowboat which is tied by a long light rope to a pier brings the boat in toward the pier by pulling on the rope. If he and the boat (and any water that goes too) weigh 400 lbs. together, and he pulls with a force of 20 lbs., what velocity does he attain in 3 sec., neglecting friction? How is this result affected if, instead of being tied to a pier, the boat is tied to another boat which weighs 800 lbs.?

13. The brakes of a 3000-lb. automobile are just strong enough to hold it at rest when it is on a hill whose vertical fall is 1 ft. for every 5 ft. along the slope. If this car moves along a level surface at the rate of 40 ft./sec., and the brakes are applied, (exerting the same force as above) in how short a distance will the car be brought to rest?

14. Find the force in a rope by means of which a 4-ton motor truck tows a disabled 1-ton automobile with an acceleration of 4 ft./sec.^{2} along a level road. Find also the total horizontal force between the driving wheels of the truck and the ground during this time.

15. A boy ties a 100-gram stone on the end of a long light string, which he coils loosely at his feet. Holding the other end in his hand, he throws the stone away from him so that it pulls the string out into a horizontal line, and at that moment the stone gives the string a powerful jerk due to the stopping of the stone's motion. If the stone is moving at the rate of 5 meters a second, and is stopped in one-tenth of a second, find the average pull on the string during the time of the jerk.

16. A stream of water is being turned upon a burning building. The water is moving horizontally at a speed of 20 ft./sec., when it strikes a wall and drops straight down from the wall after striking it. If 640 pds. of water are being delivered per second, find the steady push (in lbs.) produced by the stream on the wall.

17. Find the average force between a 2000-pd. automobile and a heavy stone wall during a collision, if the automobile was moving at the rate of 12 ft./sec. just before the collision, and its center of gravity was brought to rest in 6 inches.

18. A player kicks a football which weighs 2 lbs. and gives it a velocity of 50 ft./sec. If the contact between his foot and the ball lasts 1/50 sec., calculate the average force between his foot and the ball during this time.

19. A boy on a bicycle pulls a 50-pd. cart after him by a horizontal cord. Find the tension in the cord when he starts with an acceleration of 2 ft./sec.^{2} If he and the bicycle together weigh 100 lbs., find the horizontal force at the same time between the driving wheel of the bicycle and the ground.

20. A train is moving at the rate of 20 ft./sec. past a man running at half this speed. The man boards the train by grasping a hand-rail. If he weighs 160 lbs., and comes to rest on the train in 2 seconds, how great an average force must exist between him and the hand-rail during this time?

"Centrifugal force." There are many interesting phenomena which are met with in the consideration of rotating bodies. The first of these is furnished by the so-called centrifugal forces, or forces which seem to pull away from the center of rotation. These are found in many practical devices, such as cream separators, centrifugal driers, steam-engine governors, and centrifugal pumps. When, as occasionally happens in a mechanical shop, a rapidly spinning emery wheel flies violently to pieces, it is due to a lack of strength to withstand centrifugal force. The amount of this force under some circumstances is astonishing, and it is very important to be able to calculate in advance how large it may be in any given case.

Calculation of centrifugal force. A body revolving uniformly in a circle, as for instance a ball on the end of a string, has already (p. 56) been shown to have a uniform acceleration given by $a = v^2/r$. It must therefore be acted on by a force, according to Newton's second law, which is in the direction of the acceleration, that is, toward the center of the circle. The amount of this force is very conveniently given by the ratio equation

$$\frac{F}{W} = \frac{a}{g} = \frac{\frac{v^2}{r}}{g} = \frac{v^2}{rg}; \text{ or by } F \text{ (in dynes)} = ma = \frac{mv^2}{r}.$$

F is the force pulling the body toward the center, ~~and hence~~ called the centripetal force. According to Newton's third law there must be an equal and opposite reaction to this force, which in this case

is one of the sort we have termed (p. 67) "inertia reactions." This reaction enables the body to pull on the string (or whatever is holding it in) with a force which is called the "centrifugal force." The inertia of the body keeps it moving in a straight line, so that it tends to fly off on a tangent, thus putting a tension on the string.

The nature of centrifugal force. Though the body pulls on the string, there is no external force acting outward on the body itself. The inertia reaction is something created in the body by virtue of its acceleration. If a person revolves with the body, however, or follows the state of affairs with his mind's eye as though he did so, he sees the body acting as if it were pulled outward by some outside agency, with a force which is just balanced by the inward pull of the string. To such a person, if the string breaks, the body seems to start moving outward along a radius, and not along the tangent as it actually does. This is an effect due to his own rotation. From his standpoint, then, we may think of centrifugal force as an outward pull on the body, and this way of looking at such cases is often useful.

An analogous but simpler case of an inertia reaction, more like those already mentioned (p. 67) is that of a man sitting in a train which is moving forward with an acceleration. If he holds a weight in his hand, it appears to pull backward on him so long as the acceleration lasts. The weight does not actually try to move backward, but merely to stay at rest. The man has to push forward upon it to give it the acceleration of the car and its inertia reaction takes the form of a backward push against his hand. From the standpoint of the moving passenger it is allowable to think of this inertia reaction as an outside force acting on the weight itself, and to represent it as a backward vector. But a stationary observer would see clearly that though the weight pulls on the man's hand, there is no outside force enabling it to do so.

Effect of centrifugal force on weight. An interesting effect due to centrifugal force is the difference in the weight of objects on the earth's surface due to the fact that the earth is rotating on its axis. At the equator the velocity that bodies have on this account is 24,000 mi. (once around the earth) in 24 hr., or 1467 ft. per second. From these figures their acceleration, v^2/r, comes out very nearly 0.1 ft./sec.^{2} Since the acceleration of gravity is approximately 32 ft./sec.^{2}, we see that bodies at the equator have their accelerations, and therefore their weights, altered through the rotation of the earth by about $\frac{1}{3}$ of one per cent. Their weights would be increased in this proportion if the rotation were to cease.

At points on the earth in latitude 45°, the velocity of bodies is less than it is at the equator, since they move in a smaller circle. Moreover, as one can see by examining Fig. 5–5, the center of the circle in which they move is not the center of the earth, but a point A on the earth's axis which is well up toward the pole. The gravitational attraction of the earth on a body B is in the direction BC. The centrifugal force may be regarded as being along BF, drawn outwards since we naturally consider ourselves as rotating with the point B (as on p. 71). The effect on the weight BH is therefore to reduce it slightly, and also to tip it in a southerly direction (as shown by the dotted diagonal BG). This action is responsible for the shape of the earth, which is not perfectly round, bulging out toward the equator and flattened at the poles. The surface is nearly perpendicular to the line of the resultant force at every point.

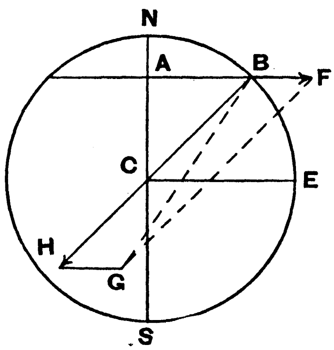


FIG. 5–5

Practical applications. A modern form of centrifugal water pump is shown in Fig. 5–6. The water is drawn into a space in which spiral vanes revolve, which impart a rotation to the mass of water, but do not retain it. It is thus driven to the outer part of the chamber and cannot return, in the absence of any centripetal force. Such pumps have no valves and are on this account better able to drive water which contains impurities or dirt.

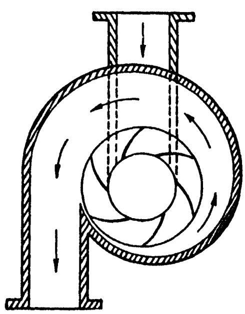


FIG. 5–6
A centrifugal pump

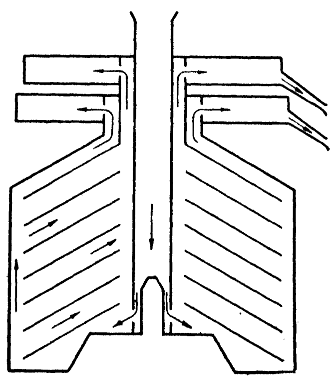


FIG. 5–7
A cream separator

The essential part of the cream separator, Fig. 5–7, consists of a cylinder rotating at an extremely high speed, often 8000 times a minute, into which the fresh milk is led. The slightly greater density of the "skimmed milk" gives it more inertia, so that it moves to the outer part of the cylinder, forcing the cream to the center. Suitable openings are provided at the center and in the outside wall to draw off the cream and the milk. The speed with which the separation occurs is very much greater than that produced in the old-fashioned

way, by gravity alone. What causes the separation in an ordinary pan is a small difference in weight, due to the difference in density of the liquids; but when rotating at high speed, the forces acting on the two liquids are greatly increased, and with them their difference is also increased, upon which the speed of separation depends. The example which follows will make this clearer. Chemists and other scientific men use "centrifuges," which accomplish the same thing.

Example on centrifugal force. To find the separating force between 1 cm.^{3} of milk (assumed to weigh 1.03 gr.) and 1 cm.^{3} of cream (0.93 gr.) when they are (*a*) in a pan under gravity, and (*b*) in a centrifugal separator at a distance of 10 cm. from the axis of rotation, and spinning at the rate of 6000 revolutions a minute.

(*a*) Under gravity the two forces are 1.03 and 0.93. Their difference, 0.1 gr., is the separating force.

(*b*) For the milk $\qquad \dfrac{F_1}{W_1} = \dfrac{v^2}{rg}$

For the cream $\qquad \dfrac{F_2}{W_2} = \dfrac{v^2}{rg}$

The velocity $\qquad v = 2\pi \times \text{(radius)} \times \text{(number of revolutions a second)}$

$$= 2\pi(10)(100) = 6280 \text{ cm./sec.}$$

The separating force is

$$F_1 - F_2 = (W_1 - W_2) \times \frac{v^2}{rg}$$
$$= (0.1)\ 4024$$
$$= 402.4 \text{ grams.}$$

The separating force is thus over 4000 times as great as in (*a*), and the process requires a second or so instead of several hours.

PROBLEMS

1. A 100-gram stone, tied to a cord whose breaking strength is 1 kg., is whirled about in a horizontal circle of 50 cm. radius at a slowly increasing rate of speed. When the cord finally breaks, with what velocity does the stone fly off?

2. A 200-gram stone is tied to the end of a string, and is then whirled in a horizontal circle of 1 m. radius at the rate of 2 revolutions a second. Find the force in the string.

3. A stone of 2-pds. mass is tied to one end of a light rope which passes through a metal ring. A man holds the ring in his hand above his head and moves it in such a way that the stone whirls in a horizontal circle of 3 ft. radius, making one revolution a second. The other end of the rope hanging from the ring is fastened to a weight which is just lifted from the floor under these conditions. How large a weight is needed to balance the centrifugal force of the stone, assuming no friction at the ring? Is the equilibrium stable, neutral, or unstable (assuming constant rate of revolution)?

4. An emery wheel in a machine shop has a diameter of 20 cm., and whirls at the rate of 50 revolutions per second. What is the centrifugal force on each gram of material on its outer edge?

5. An aviator is doing a vertical loop at the top of which he is upside down. If he is then tracing part of a circle of 200 ft. radius, what must be the slowest speed at which he can make the loop and still press against his seat in the direction which seems usual to him?

6. At an amusement resort there is a framework CD (Fig. 5-8) which revolves around a central pole AB and supports carriages E and F on chains in which people may sit while the framework revolves. BC is 10 ft. long; EC is 16 ft. It is desired to have the chains CE and DF stand out no farther than 45°. What is the greatest speed allowable, and how many revolutions per minute does this involve?

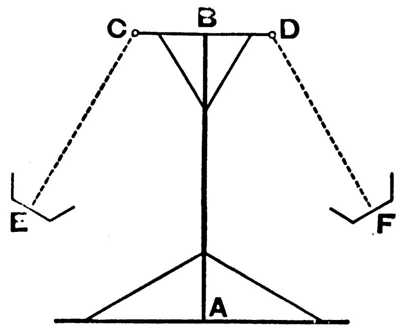


Fig. 5–8

7. A motorcycle is racing around a circular track of 2000 ft. diameter at a speed of 100 ft./sec. Find the centrifugal acceleration and show by means of a drawing the angle at which the rider must lean.

8. How much should the outer rail of a track 56 in. wide be elevated on a curve of 2000 ft. radius where trains are supposed to maintain a speed of 70 ft./sec.?

9. A 4000-lb. racing automobile runs due east along a level track at the equator at the rate of 120 miles an hour (176 ft./sec.). The ground there is carried eastward at the rate of 1500 ft./sec. on account of the rotation of the earth. By how much does the motion of the car alter its weight due to the change in the centrifugal force? (Answer 3.3 lbs.).

10. An automobile rounds a sharp curve at a speed of 15 ft./sec. A 150-lb. passenger is thrown against the side of his seat. With what force does he press against it if he is traveling on a circle of 25 ft. radius?

11. A belt connects an engine with a large machine, and in passing over the pulley of the machine (at rest) exerts a pressure of 1 lb./in.^{2} The belt has a mass of 0.07 pd./in.^{2}; its usual speed is 10 ft./sec.; the diameter of the pulley is 2 ft. To what value is the pressure between the belt and the pulley reduced by centrifugal force when the machine is running? (Answer 0.78 lb./in.^{2}).

FRICTION

Causes of sliding friction. When two bodies are pressed against each other, they make contact in spots, due to uneven surfaces, and it requires some force to rub one over the other, either because

the protruding particles are being torn away, or because one body has to be lifted over the little hills in the other, as it were, in order to make progress. This we call the force of friction. It appears as a destructive rather than a creative agent. By its means we lose power and waste energy in all machinery, causing trouble and expense; but to it we owe the possibility of making buildings, clothing, etc., and of moving over the surface of the earth; so that, on the whole, it does much more good than harm.

Rolling friction. When an automobile wheel is rolling over a hard level road, the part of the tire which is in contact with the road is temporarily flattened; so that as the wheel travels each part is flattened in succession, and then recovers its shape again. In the course of this action there is internal friction, due to disarrangement of the parts of the material. No substance is perfectly elastic. Even in the case of steel such actions will eventually "wear out" the material, and the internal forces will cease to hold it together. This type of friction is called "rolling friction" and can be made very small in machinery by the use of steel ball or roller bearings, though it cannot be entirely eliminated. Even in a ball-bearing a hard steel ball is deformed as it rolls around, in just the same manner as the automobile tire, though to a much smaller extent. Eventually it will get "old" and is then likely to break. An interesting example of such deformation is furnished by the experiment of dropping a hard round ball, such as a billiard ball, on a smooth iron surface previously blackened with soot from a candle flame. The ball will rebound from the iron, but will carry with it a round disc of compressed lampblack covering the area of contact. This area will be larger the greater the height of fall, and it indicates an amount of yielding at the point of contact which is usually unexpectedly large.

Coefficients of friction. Sliding or "kinetic" friction, while somewhat variable, follows approximately a simple relation. If F is the force dragging an object over a surface, and P is the force pressing the object and the surface together, the ratio F/P is nearly constant, and is known as the ***coefficient of friction.*** There is, however, one amount of force required to ***start*** the motion, and another smaller amount needed to maintain it after it is started. This leads to ***two*** coefficients called the ***coefficient of starting friction*** and the ***coefficient of sliding friction.*** Measurements of these quantities are often made in the laboratory, or on the lecture

table, and are chiefly useful in showing how very sensitive the force is to the least film of foreign matter on the surface, or to minute differences of polish not otherwise detectable.

Lubrication. Viscosity. In practice, of course, it is customary to lubricate surfaces that must slide over one another. Oil consists of particles (molecules) that are known to have an elongated shape, one end being active and likely to cling to other materials (see p. 132), while the opposite end is inactive and slips readily over any other particles with which it comes in contact. An oil film will thus cling to a metal surface, and when two oiled metals rub over each other, it is really a film of oil which rubs over another film of oil, the metals not coming into contact at all. If however they are pressed too firmly together the oil may be squeezed out entirely, and then the friction increases so much that a great development of heat occurs, and we have a "hot-box" or a "burned-out" (i.e. melted out) bearing. A thick, slow-running oil is less likely to be squeezed out of a bearing since it has a high viscosity; that is, there is a considerable amount of friction connected with a flow of one layer of the liquid over another. The best arrangement for lubrication is to have the oil continually drawn in to a narrowing space as the motion progresses. For instance, in the case of a shaft rotating in its bearing, the shaft does not quite fill the space in which it rotates (Fig. 5-9), and the advancing surface of the shaft will tend to entrap oil in the narrowing space between it and the wall of the bearing. Under ordinary running conditions the amount of friction depends on the viscosity and also varies directly with the speed of one surface with respect to the other. If the speed is sufficient, it will make up for a low viscosity. Cream separators can be lubricated with milk, and several machines use an air film as the only lubricant.

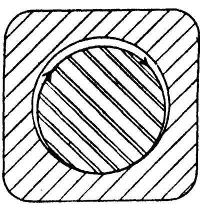


Fig. 5-9

Dimensions of physical quantities. There is a way of distinguishing physical quantities, such as force, momentum, etc., which is simple, interesting, and useful. This is by means of what are called their *dimensions*. The method can best be shown by examples.

A length is completely specified by one number and the name of the unit used. Only a single unit is involved, which we might designate simply by L. In the language of this subject we say that a *length* has the dimension L. Likewise a *time* has the dimension T and we may say that a *mass* has the dimension M. Having chosen these three, all other mechanical quantities can be expressed in terms of them. A *velocity* has the dimensions L/T, being made up of some length divided by an amount of time. An *acceleration* must be specified by L/T^2, since it is a change of velocity divided by a time. A *force*, according to the second law, is a mass multiplied by an acceleration; it has the dimensions ML/T^2. A *momentum* is characterized by ML/T.

These conventional representations may be very helpful in avoiding mistakes. For example, every equation must have the same dimensions on both sides. If a student wishes to use the formula for centrifugal force but has

forgotten whether it is mv/r or mv^2/r, he can see at once that mv/r gives the dimensions M/T (the length in the velocity canceling the length in the radius) and therefore is not a force. If he has forgotten everything about the formula he might reason it out by dimensions. Thus he might assume from common-sense grounds, that the centrifugal force must depend somehow on the mass, the velocity and the distance. If he then assumed it to be

$$M^x \times v^y \times r^z \quad \text{or} \quad M^x \times \frac{L^y}{T^y} \times L^z \quad \text{or} \quad M^x \times \frac{L^{y+z}}{T^y},$$

he should know that this was equal to ML/T^2, the dimensions of any force. Hence, by equating coefficients, x must be unity to make the M's agree; y must be 2, to make the T's agree, and z must be -1 to make the L's agree, and the result is mv^2/r. Naturally, such a dimensional proof does not indicate whether there are any constants in the formula or not.

Other uses of dimensional reasoning are in working out the behavior of airplanes, or ships, from laboratory measurements on small-scale models. Or, as Haldane [1] has done, the necessary thickness of a giant's legs can be calculated when his height is known.

PROBLEMS

1. A 5-kg. stone is thrown out on level ice with a velocity of 2 m. a second If it comes to rest in 20 m. distance, find the average frictional force opposing it, and the coefficient of friction.

2. A child is pulling a heavy toy about on the floor at a uniform speed by means of a string which makes an angle of 60° with the floor. If the force in the string is 2 lbs., what is the horizontal force acting on the toy, and how great a frictional force must there be between it and the floor?

3. A boy sits upon a level, circular platform at a distance of 2 ft. from its center. The coefficient of friction between him and the platform is 0.1, and his weight is 100 lbs. What is the greatest force that friction can furnish to keep him from sliding off when the platform begins to revolve? What rate of rotation must the platform attain so that he can no longer hold his place?

4. A 50-kg. sled slides down a hill whose slant distance is 50 m. and vertical height 10 m. If it acquires a speed of 10 m. a second at the bottom, find the average force of friction (in kg.) opposing the motion during the descent.

5. A 100-lb. sled slides down a hill whose vertical height is one-tenth of its length along the slope. If the force of friction between it and the surface is 2 lbs., find the velocity acquired by the sled after it has gone 100 ft. down the hill.

6. A 2000-lb. automobile is "coasting" down a hill at the constant speed of 40 ft. a second. If the hill descends 1 ft. for every 10 ft. along the slope, how much frictional force must be resisting the motion?

---

[1] J. B. S. Haldane, English physiologist. See his delightful essay "On being the right size" in *Possible Worlds*, 1928 (Harper and Brothers).

7. How much force would the driving wheels of the automobile in the preceding problem have to exert parallel to the road in order to keep the car *ascending* the same hill at the same speed, assuming the same friction? (Consider only what is needed after the car has reached this speed.)

8. An automobile weighing 2600 lbs. is traveling at a speed of 45 ft./sec. around a curve whose radius is 200 ft. Find the horizontal force which is needed to prevent skidding, and the least coefficient of friction between the roadway and the tires. Assume the road not banked.

9. If a block of wood is resting on a level board one end of which can be raised, it will begin to move when the board makes a certain angle with the horizontal. Show that the tangent of this angle is equal to the coefficient of friction.

10. The coefficient of friction between a 48-lb. box and the floor is 0.1. Find the force required to drag the box over the floor (*a*) at a uniform speed, (*b*) with an acceleration of 2 ft./sec.^{2}

11. Each driving wheel of an automobile carries a weight of 600 lbs. What must be the least coefficient of friction between the wheel and the road in order that each wheel may be able to push horizontally with a force of 100 lbs.?

12. An automobile has to tow a heavily loaded wagon along a slippery road. If the coefficient of friction between the wheels of the automobile and the road is 0.15, and a 150-lb. pull is needed to keep the wagon going, how heavy must the automobile be, assuming its weight to be borne equally by its four wheels?

13. Test the formulæ for falling bodies (p. 46) from the standpoint of dimensions.

14. What are the dimensions of pressure, of volume, of area, and of density? If (on p. 6) it had been said that $p = hd$, what criticism could have been raised against this statement on dimensional grounds?

15. Make up a dimensional system on the assumption that the three fundamental quantities are length, time, and force. (This system has been used.)

16. A card rests under a pile of papers. It can be pulled out by a jerk without moving the papers. Why?

17. A straight stick rests horizontally on two supports (e.g. fingers) near its ends. One support is then moved slowly toward the other, keeping the stick horizontal. Explain why the stick does not fall. (Try it.) Why must the motion be slow?

CHAPTER 6

WORK, ENERGY AND POWER

The technical terms used in mechanics are frequently words that are used also in common speech in a variety of senses. Learning the precise meanings to which these terms are restricted in their scientific use is one of the difficulties of the subject. Nowhere is this more marked than in the consideration of work and power; nor are there in this mechanical age many terms which are more frequently used.

Nature of work. Work is formally stated to be the product of a force and the distance through which the force succeeds in moving the object acted upon. It is understood that the distance is always measured along the direction in which the force acts, which will not in all cases be the direction in which the body is moving. This is a sensible definition. If one wants coal hauled out of a mine, or water pumped into a high reservoir, the weight lifted and the height raised each enter in, and the product of the two gives a proper measure of the work done. Horizontal motion on the part of the water, or the coal, does not count, if we neglect friction, and it should not, since it contributes nothing to the height. Similarly, a man holding a heavy suitcase is doing no work, even though he is tiring his muscles, because he is moving nothing. If we should credit him with working under the circumstances we should also be driven into the absurdity of crediting his bed with doing a large amount of work because it held him off the floor while he slept.

Work units. The units of work are various, as we may combine ***any*** force unit with ***any*** distance unit to create a work unit. We shall restrict ourselves to a few of these combinations only.

If the force of one pound acts through a distance of one foot, one ***foot-pound*** of work is done; if a gram of force is exerted through a centimeter, one ***gram-centimeter*** of work is done; if a dyne of force acts through a centimeter, one dyne-centimeter of work, called an ***erg***, is accomplished. The last unit is the only one of these three which has received a name of its own, but since the dyne is a very small unit, the erg is usually inconvenient; more often in practical problems we use the ***joule***, which is equal to 10,000,000 ergs.

All units of work have the ***dimensions*** of force $\times$ distance, or $ML/T^2 \times L$, which is ML^2/T^2. It is to be noted that these are also the dimensions of a mass multiplied by the square of a velocity, which we shall see is kinetic energy.

Simple machines. Simple machines are devices for doing work and at the same time altering the amount or direction of the force needed to do it. A crowbar, used to pry up a heavy stone, is a familiar example. The force F (Fig. 6–1) which must be exerted on the handle overcomes the resistance R, with a negligible amount of help from the weight of the crowbar. As the fulcrum is a long way from the handle, the torque produced by the force F is large. The second principle of statics (p. 33) states that the torques produced by all the forces must balance, and this serves as a basis for calculations. One can see that the force R is much greater than F; their ratio is called the ***mechanical advantage.***

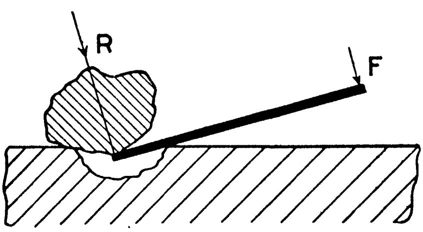


Fig. 6–1

The ***efficiency*** of any machine is the work accomplished divided by the work put into the machine. In the case of the lever, it is nearly 100%; but in many common forms of machines it is quite low. Obviously it is a quantity of great practical importance.

A set of ***pulleys*** as in Fig. 6–2, shows another simple way of securing a large mechanical advantage, and lifting heavy loads by means of small forces. The load is equally divided among all the cords supporting it, and the man working the hoist has only to overcome friction and supply one-fourth (in this case) of the force needed to support the load. The work done by his hand depends on the distance through which it moves. Evidently in

this example it moves four times as fast as the load does. Hence, if a force one-fourth as large has to move through a distance four times as great, the work done is the same. As a matter of fact the hand-force must be larger in order to overcome friction, so that an amount of extra work has to be done; but this may not always be considered important if by means of the machine weights can be lifted which would otherwise be unmanageable.

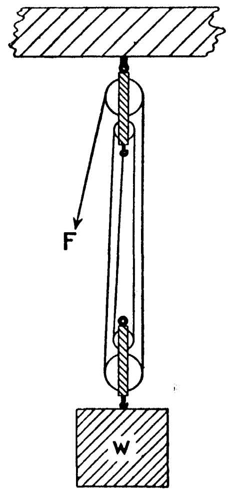


FIG. 6–2

The force required to pull a body up an ***inclined plane*** is small compared with the weight of the body. Hence the inclined plane should also be classed among the simple machines. There are many other types of machines which make interesting objects for laboratory study.

Example on work. A loaded motor truck weighs 3 tons. It climbs a hill 600 ft. long rising 1 ft. vertically for each 20 ft. of slope, and a force of 50 lbs. due to friction opposes the motion. Find the work done.

The work done against gravity is that due to lifting 6000 lbs. up a slope involving 30 ft. of vertical rise; it is therefore 6000 × 30, or 180,000 ft.-lbs. The work done against friction is the force along the slope due to friction multiplied by the distance along the slope, or 50 × 600 = 30,000 ft.-lbs. The total work done is therefore 210,000 ft.-lbs. The efficiency of the arrangement is evidently 18/21, or 86%.

Power. An agent which continues to produce work is said to exert power, and this power is ***measured by the rate of doing work.*** Of the many possible units of power only two are much used; the ***horse-power,*** which does work at the rate of 550 ft.-lbs. per second, and the ***watt,*** which works at the rate of one joule per second. For industrial purposes the ***kilowatt*** (1000 watts) is commonly found more convenient than the watt. It is also sometimes useful to know that 746 watts are equal to 1 H. P. In commercial practice the kilowatt hour is used as a unit of work; the equally awkward term horse-power hour is also often encountered.

Sources of power. Power is commonly derived from burning coal, oil, or gasoline (see engines, Chapter 15) or from falling water. In the latter case ***turbine wheels*** are used. The water is usually allowed to fall in a smooth pipe through most of the available height, and thus to gather speed. It then impinges on a set

of vanes set around the wheel, somewhat after the fashion of a windmill, or of the paper pin wheels that children make. These vanes deflect the water and bring it nearly to rest, thus depriving it of most of its kinetic energy, and deriving the greatest possible power from it. The resulting force driving the wheel depends on the amount of momentum destroyed per second (Newton's second law). In order to destroy as much as possible, the best angle and shape of the vanes for different speeds have been most carefully studied. Figure 6–3 shows the type of vanes used in a power plant of recent construction. A somewhat similar problem is found in the case of steam turbines (p. 213), and in airplane propellers.

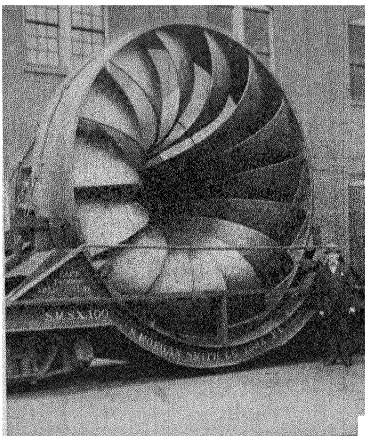


FIG. 6–3
(Courtesy of the S. Morgan Smith Co.)

Example on power. A waterfall delivers 50 cu. ft. of water each second over a fall of 20 ft. of vertical height. If the necessary machinery to convert this into useful power involves a loss of 20 per cent through friction, etc., find the total power obtainable from this fall, expressed both in horse-power and in kilowatts.

Fifty cu. ft. of water weighs 50×62.4 or 3120 lbs. The work done by this weight in falling 20 ft. is 62,400 ft.-lbs. Eighty per cent of this is saved, or 49,920 ft.-lbs., and this work being produced per second gives a rate of 49,920/550 or 90.8 horse-power. This is equivalent to 90.8×746 watts, or 67.7 kilowatts.

Energy. A body is said to possess energy because it can do work, and the energy is measured by the work that can be obtained from it under the circumstances in which it is placed. Ordinarily, energy can be in two forms, potential and kinetic.

Potential energy is energy due to advantageous position or configuration, as in the case of a raised weight, a bent bow, a wound-up clock spring, etc. This is often easily measurable. For instance, in the case of the raised weight, the potential energy is the weight multiplied by the height through which it can fall. The potential energy of a stretched spring is equal to the ***average*** force required to stretch it (half the maximum force) multiplied by the distance of stretch. The potential energy of a body may sometimes be

altered by changing the external circumstances, without touching the body; for instance, the energy of a raised weight may be changed by boring a hole in the floor directly under it, so that it can fall farther than before. This makes the amount of potential energy in a body indefinite unless the circumstances are exactly specified, as, of course, they usually are.

The ***kinetic energy*** of a body is the energy due to its motion. A moving body may be made to do work, to compress a spring by hitting it, to raise weights, or to start another body moving. Kinetic energy is frequently being transformed into potential energy, or conversely, as the body moves. It is interesting to consider the energy changes going on in the case of a simple pendulum, made up of a heavy weight hung on a fine wire and started swinging. For a long time its swings may not diminish appreciably, but in every swing its energy is transformed from kinetic in the middle, when it is as low as possible, to potential at the ends, when it is as high as it will go, and is momentarily at rest. Evidently if its swings persist, its total energy remains unchanged, and the transformation from one form to the other is accomplished without loss. Eventually, of course, such a pendulum will come to rest, the energy being "lost" by friction. We know that it is thus transformed into a third form of energy, heat, a form less easily observed, but which we have reason to believe is a combination of kinetic and potential energies.

Conservation of energy. A pendulum, as above, furnishes a simple example of a principle of great importance, the principle of the ***conservation of energy.*** This principle states that ***the total amount of energy possessed by an isolated body, or system of bodies, remains constant.*** It may change from one form to another, or pass from one body to another inside the system without loss. In practice, it is not always easy to isolate a body so that energy cannot leave it, or get into it from its surroundings; but the most careful measurements have as yet shown no exceptions to the principle, when allowances have been made for the effects of friction, escape of heat, etc.[1] Nor have the machines devised to run with perpetual motion, in defiance of the principle, ever lived up to the hopes of their inventors.

[1] We cannot yet be quite certain that the principle of the conservation of energy holds with reference to processes that take place within an atom, but it is certainly valid for pieces of matter large enough to be seen and handled.

Perpetual motion machines. Many of these are of great ingenuity and seem workable at first sight. Most of them, however, are devised by simple-minded people who think that falling water can be made to run a pump, which will in its turn raise the water all back to its original height, or perhaps a little higher. Slightly less crude is a plan such as this Fig. 6–4: — a chain runs over three cogwheels, two of them in a vertical line and one to one side. The extra weight of the chain on one side is supposed to pull that side down. The fallacy is not hard to discover.

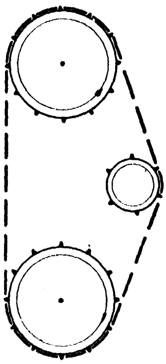


Fig. 6–4

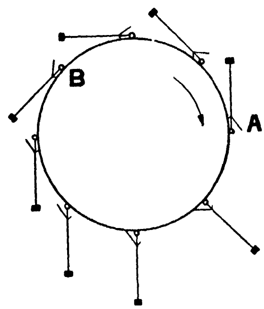


Fig. 6–5

A more interesting mechanical form (Fig. 6–5) is this: — a wheel has a number of short rods hinged at points regularly placed around its rim, each carrying a weight at its tip. Each rod is free to turn through a right angle, between stops, from a radial to a tangential position. Assuming rotation in the sense indicated, the rod at A is evidently on the point of falling down and becoming radial, while those near A are all tangential. In the radial position the weights have greater torques and might be supposed to be able to produce rotation. A little consideration of the torques produced by the other weights will show why the wheel does not go.

A more elaborate and quite different idea was to manufacture liquid air and then use it as a motive power, presumably with the expectation of getting back at least as much power as was put into the manufacturing of it. This last example led to the investment of a quantity of good money in a bad cause not so very many years ago. Even at the present time, more than seventy years after the first statement of the principle of the conservation of energy, the newspapers sometimes contain accounts of inventions by which it is hoped to obtain large amounts of power out of nothing.

Calculation of kinetic energy. It will be very useful to know the amount of energy possessed by a body of mass m when it is moving with velocity v. Suppose it to have reached this state by being pulled by a force F, which started it into motion (without friction) and acted steadily on it until it acquired the velocity v. The force would have done a quantity of work Fd on the body, where d is the distance it went in the time t during which the motion occurred. It would move with uniform acceleration a, and the velocity it acquired would be given by the equation $v^2 = 2ad$; hence $a = v^2/2d$. If the force is measured in absolute units (dynes, for instance)

$$F = ma = m \times v^2/2d, \text{ whence } Fd = \tfrac{1}{2}mv^2,$$

or the work done, in absolute units, equals half the mass multiplied by the square of the velocity. By the principle of the conservation of energy, the work done is equal to the kinetic energy acquired, since the motion is supposed to be along a level plane. Hence the kinetic energy (K.E.) $= \frac{1}{2}mv^2$ in ergs, or other absolute units. If the force is expressed in gravitational units (pounds or grams)

$$F/W = a/g \quad \text{or} \quad Fd = \tfrac{1}{2} \times W/g \times v^2$$

using the equation $a = v^2/2d$. This expression gives the kinetic energy in gravitational units, such as ft.-lbs. or gr.-cm. The ***dimensions*** of energy are, of course, the same as those of work.

Examples on energy. (*a*) Find the kinetic energy of a 2500-lb. automobile, moving with a velocity of 45 miles an hour (66 ft. a sec.).

K.E. $= \frac{1}{2}.\ 2500/32 \times (66)^2$ ft.-lbs. $=$ 170,000 ft.-lbs. approximately.

(*b*) Find the average horse-power required on the part of the engine of this automobile to produce this speed in 20 sec., neglecting friction.

170,000/20 = 8500 ft.-lbs. produced per second, requiring
8500/550 = 15.5 horse-power.

(*c*) Find the kinetic energy in joules of a rifle bullet of 2-gm. mass moving with a velocity of 300 meters a second.

K.E. $= \frac{1}{2} \times 2 \times (30{,}000)^2 =$ 900,000,000 ergs = 90 joules

(*d*) Find the power used by a fly whose weight is 0.1 gr. in climbing vertically up a wall at the rate of 5 mm. a sec.

It does $0.1 \times 0.5 = 0.05$ gram-centimeters of work per second, or $980 \times 0.05 = 49$ ergs per second. The answer may well be left in these units, since small units are appropriate in this case.

The conservation of energy as an aid in solution of problems. Such problems as those dealing with collisions may often be very conveniently solved by the use of the principle of the conservation of energy. As an example consider the following case: — A pile driver consists of a 50-kg. weight, raised to a height of 8 meters and let drop on the top of the pile. Assuming that no energy is lost in such forms as heat, sound, etc., which would be approximately true in practice, suppose that the pile is driven in a distance (d) = 25 cm., find the average force exerted on the pile during its motion. The work done is Fd, which is equal to the kinetic energy of the weight just before it strikes the pile, and this in turn is equal to the potential energy of the weight in its raised position, or the work done in raising it, which is Wh if W is the weight and h the height. Hence $Fd = Wh$, or $F = 50$ kg. $\times 800$ cm./25 cm. $= 1600$ kg.

Example. An automobile weighing 2500 lbs. and moving with a speed of 20 ft./sec., crashes into an immovable stone wall, and its center of gravity moves forward 1 ft. during the collision, before it is brought to rest. Find the average force between the car and the wall, and the time of duration of the impact, assuming uniformly accelerated motion during this time.

The kinetic energy of the car is $\frac{1}{2}mv^2$, which is $\frac{1}{2} \times 2500 \times (20)^2$ in foot-poundals, or 1/32 as much in ft.-lbs. This gives 15,625 ft.-lbs. This is the work done by the force we seek acting through 1 ft. The force must then have been 15,625 lbs. The duration of the impact is most easily found by recalling the formula $s = \frac{1}{2}vt$ for uniformly accelerated motion. Here we have negative acceleration, but this merely means that the acceleration is in the opposite direction to the initial speed. Since s, the distance traveled, is 1 ft., it follows that $t = 2s/v = 2/20 = 0.1$ second.

MOVING LIQUIDS

New effects introduced by motion in liquids. Certain curious effects are encountered with moving liquids that should be considered, since they are often of great practical importance. The laws of hydrostatic pressure no longer apply when motion occurs. The pressure formulæ need modification, for instance, in the case of a city water system, in which the water is continually flowing; and very odd cases arise sometimes in which we have to appeal to Newton's laws of motion or to energy considerations for an explanation. Some of the conclusions reached from the study of moving liquids apply also to moving gases, though generally with modifications due to the compressibility of the latter.

Flow of liquid from an orifice. If a pail is filled with water, but has a hole in it, a stream will issue from the hole, starting in a direction perpendicular to the surface in which the hole is situated and flowing with a speed which is easily calculated if we neglect friction. A mass m flowing out with velocity v carries away a quantity of energy equal to $mv^2/2$. As a result the surface of the liquid falls, losing an equal amount of energy. It is as though the top layer of the liquid had fallen through a height h from its average level above to the level of the hole below. If m grams do so, they lose an amount of energy equal to mgh ergs; therefore, without friction, $mgh = \frac{1}{2} \times mv^2$ or the velocity of flow is $v = \sqrt{2gh}$, which is the same as the velocity of a freely falling body which has fallen through the same height. As a matter of fact, the issuing liquid flows in "lines of flow," which are not all perpendicular to the wall surface at the hole, but have an inward component, toward the center of the hole, forcing the liquid stream to contract after it escapes, and compressing it a trifle so that it continues to be accelerated for a short distance after it leaves the hole. The narrowest part of the stream is thus outside the hole, and is called the "vena contracta." The velocity of efflux can be increased by shaping the opening into a nozzle, ***following the stream lines.*** Somewhat similar considerations are met with in the motion of solids through fluids, as in the case of ships moving through water, or of airplanes through air. The resistance encountered may be very greatly reduced by shaping the solid so that the flow follows stream lines, instead of forming eddies as it otherwise tends to do, which inevitably produce energy losses and reduced speed.

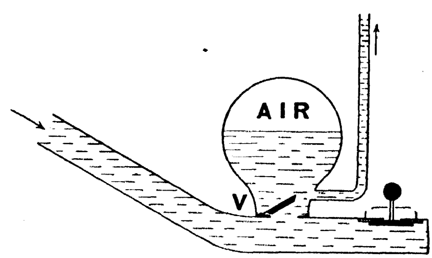


FIG. 6–6
The hydraulic ram

The hydraulic ram. It is possible for a flowing liquid to raise part of itself to a height much higher than its own surface. This is accomplished in the hydraulic ram. Water flows through a pipe, to a valve, which has a spring or weight tending to keep it open. A rush of water through this will drive it shut and thus suddenly stop the flow of a considerable column of water. This produces a large collision force, as a large amount of momentum is suddenly destroyed (Newton's second law), and this force is capable of driving a small part of the water up through the valve V to a great height, as in Fig. 6–6. The

moment the flow is stopped the valve opens again, the flow recommences, is again stopped, and so on. Usually a reservoir is furnished which is initially full of air; this becomes highly compressed, and thus serves to make the flow up the pipe more uniform.

Bernouilli's principle. An important and paradoxical action occurs in a moving fluid if it is forced to pass through a restricted region, such as a narrow piece of pipe, as in Fig. 6–7. If a small mass of water M is flowing at a moderate speed in a wide pipe and then enters a constricted portion, it must be accelerated on entering, and slowed down again to the original speed on leaving the narrow portion. The flow is continuous. All the water that flows through the wide pipe in a second must somehow get through the narrow part too, which it can do only by going faster. But the small mass M must have had a force behind it to accelerate it on entering the constriction; hence the pressure in the wide part of the tube must be greater than that in the narrow. If pressure gauges are attached at P and Q, this change of pressure will show readily. A single gauge may be arranged to show the difference of pressure at P and Q and as this difference depends on the rate of flow (though not directly) the gauge can be graduated to read rates of flow rather than pressures, and can even be made to record, as in city systems, the amount of water used in any given time. Such an instrument is then known as a Venturi meter. The fact that pressure is reduced in constrictions is known as ***Bernouilli's*** [1] ***principle;*** we omit the mathematical statement of it.

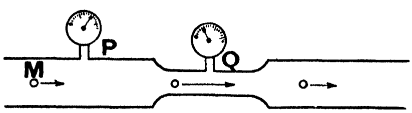


Fig. 6–7

Experimental illustrations. Many curious phenomena arise from this cause. If air is blown at considerable pressure from a nozzle provided with a wide flat end, (Fig. 6–8 shows a glass tube passing through a wide cork) and the blast is directed vertically downward on a card, it will blow the card away if the distance is great, but if it is small so that the air flow is constricted, the pressure between the card and the cork becomes so much reduced

[1] Daniel Bernouilli, (1700–1782), a member of a Dutch-Swiss family distinguished for several generations by its scientific attainments; physician, and then in succession professor of mathematics, anatomy and botany, and experimental and speculative philosophy. In his work on hydrodynamics he laid the foundation of the kinetic theory of gases.

that the card is actually drawn up almost to the cork and is held there. The swifter the air current the more the card is attracted ***toward*** it.

Two tennis balls hung on long threads near together will be drawn closer together if a blast of air is blown between them. The stream is constricted in the space between the balls, and the pressure sufficiently lowered thereby to produce an apparent attraction. Two ships anchored side by side in a river are likewise drawn to one another; or, since relative motion is all that counts in such cases, two ships steaming side by side through still water experience the same force, and collisions have occurred on this account. If a light hollow ball, such as a ping-pong ball, is

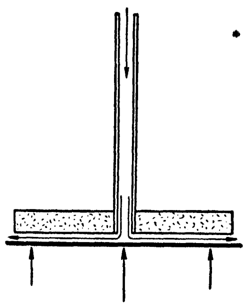


FIG. 6–8

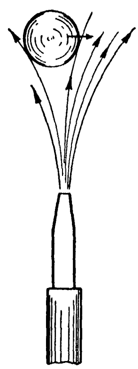


FIG. 6–9

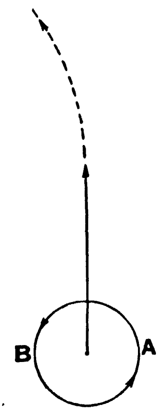


FIG. 6–10

A ball supported on a blast of air

placed in a strong concentrated upward blast of air coming from a nozzle (Fig. 6–9), it rests in stable equilibrium, supported in the middle of the air-stream by the impact of the moving air. In this case if the ball moves a little to one side of the center of the stream, the current rushing by is bent, and in a sense constricted between the ball on one side and the mass of stationary air on the other. The air pressure is considerably reduced in the main stream next to the ball, and less so on the other side. Hence the ball is pushed back into the middle of the stream, instead of being blown away. A "cut" tennis ball and a "sliced" golf ball offer similar examples. Here the air is at rest, but as before, relative motion is what counts. On account of the spin of the ball the motion of the air past the ball is greatest on the side of the ball B, Fig. 6–10,

since here the spin hurries the air past the ball at a greater rate than if the ball were not spinning. Hence on the side B the pressure is low, and the ball is deflected to the left if the direction of spin is as drawn in the figure. A somewhat similar rush of air over the top surface of the wings of an airplane produces a reduction of pressure above them, which gives a large lifting effect.

An L-shaped tube with open ends may be placed with its lower leg horizontal under the surface of a river, and pointing upstream. If so, the momentum of the current will force the water in the vertical leg of the tube to a higher level than that outside. This is one form of "Pitot tube," which may be used to measure rates of flow. If two tubes are mounted in an airplane with their open ends facing forward, the pressure in each will be increased in this way. But, if one of these is connected to a double funnel, as F in Fig. 6–11, the rush of air in the constriction will lower the pressure there, and a differential pressure gauge G will show the difference of pressure in the two tubes. This difference varies with the air-speed, and this ***air-speed gauge*** may be graduated to read directly in terms of miles per hour.

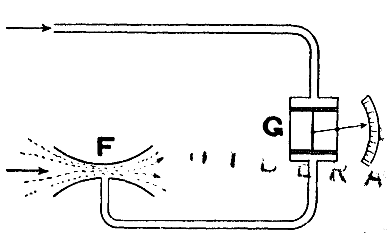


FIG. 6–11
An air-speed gauge

PROBLEMS

1. A 2-ton truck has its speed increased from 20 ft./sec. to 30 ft./sec. in 5 sec. Find the accelerating force, assumed constant, and the work done by this force in this time.

2. A 2000-gm. hammer, moving with a velocity of 2 m. a second strikes a nail and drives it in a distance of 5 mm. Find (by energy) the average force of resistance during the 5 mm. motion, and the time of the collision.

3. A football player, whose weight is 180 lbs., starts from rest with uniform acceleration and attains a speed of 8 ft./sec. in half a second. Find the average horizontal force he exerts against the ground during this time, and the average horse-power at which he is working.

4. A 3000-lb. automobile can keep up a speed of 40 ft./sec. up a hill which rises 1 ft. vertically in every 20 ft. of slope. Find the horse-power of the engine (a) neglecting friction, (b) allowing for a frictional force which uses up half the power required in (a).

5. A locomotive working at the rate of 50 horse-power pulls a train over a distance of 2000 ft. in a minute. Find the average force exerted by the locomotive during this time, assuming uniform speed.

6. A water tower 100 ft. in height has a cross-sectional area of 50 sq. ft., and is to be filled with water from a large lake 40 ft. below the bottom of the tower. How long will it take a 10-H.P. pump whose efficiency is 60 per cent to fill the tower?

7. A bullet of 2-gm. mass is fired horizontally into a 4000-gm. pendulum bob and causes this to swing back through such a distance that the bob rises 5 mm. vertically on the way. Find the velocity of the bullet before the collision, and show what energy changes occur during the collision.

8. A plank, 16 ft. long, lies on the floor. To hold one end just off the floor requires a vertical force of 50 lbs. Find the work done in lifting the plank to a vertical position.

9. A man is turning a crank at the rate of 2 revolutions per second. If the crank is 1.5 ft. long, and he exerts a steady force of 5 lbs. perpendicularly to the handle, how much work does he do in 1 minute?

10. A 160-lb. man runs uphill at a uniform rate of 10 ft./sec. measured along the slope. The hill is so steep that 100 ft. along the slope involve 20 ft. vertical rise. Find the horse-power at which he is working.

11. If the man in the last problem runs up this slope starting from rest with an acceleration of 4 ft./sec.^{2}, at what average horse-power will he be working during the first 2 sec.?

12. A steam engine drives a machine by means of a belt. The part of the belt approaching the engine is under a tension of 50 lbs.; the receding part has a tension of 10 lbs. The pulley on the machine is 1 ft. in diameter. Find the torque exerted by the engine on the machine and the power being consumed by it if the machine is running at the rate of 300 revolutions per minute.

13. An eight-oared crew makes 30 strokes per minute. If the pull in each stroke is through a distance of 4 ft. and the average pull per man during each stroke is 70 lbs., find the horse-power of the crew.

14. An elevator weighing 1 ton starts from rest and rises with an acceleration of 6.4 ft./sec.^{2} Find the tension in the cable which supports the elevator, and the work done in the first second.

15. An airplane strut offers a head resistance of 3 lbs. at a speed of 120 mi./hr. What H.P. is taken from the engine at that speed to overcome the drag of the strut?

16. The engine of a one-ton airplane stops while it is flying horizontally at a speed of 176 ft./sec. at an altitude of 4000 ft. The plane descends without power to the ground, making a landing at 80 ft./sec. What average wind resistance did the plane meet in the descent, assuming that it traveled just 5 mi. during the process?

17. A system of pulleys is so arranged that the end of the rope where a force of 80 lbs. is applied moves through 12 ft. while the load of 350 lbs. moves through 2 ft. Find the mechanical advantage and the efficiency.

18. A horse has to exert a force of 100 lbs. to pull a wagon weighing 1000 lbs. (with its load) up a hill whose length along the slope is 300 ft. If the vertical height of the hill is 20 ft., find the work done by the horse, the work accomplished against gravity, the mechanical advantage of the hill considered as a machine, and the efficiency of the arrangement.

19. Find the work done in raising a stone of 150-pds. mass and 1 cu. ft. volume through 10 ft. vertically, all of this distance being under water.

CHAPTER 7

ROTATION [1]

Rotating bodies are common in daily experience, examples being furnished in nearly all sorts of machinery. Such bodies move according to mechanical laws very similar to those followed by bodies moving in straight lines. The differences are due to the peculiarities of rotary motion which require the use of a few new terms. For convenience we shall consider these first.

Angles can be measured in degrees, or in ***circular measure;*** the latter is the more convenient system to use in equations describing rotary motion. The unit of angle in circular measure is the ***radian,*** an angle whose arc is equal to its radius. In general (Fig. 7–1) if s is the arc of an angle θ (theta),[2] $\theta = s/r$ radians, or $s = r\theta$. If a point is moving uniformly over this arc with speed v, its radius will also be describing an increasing angle at a rate called the ***angular velocity.*** Just as s/t gives v, so θ/t gives the angular velocity, called ω (omega). Since $s = r\theta$, $s/t = r\theta/t$ or $v = r\omega$. If the moving point is being uniformly accelerated around the circle, $a = v/t = r\omega/t = r\alpha$, where $\omega/t = \alpha$ (alpha), the ***angular acceleration,*** which is measured in terms of radians per second per second, in analogy with ordinary acceleration in ft./sec.^{2}

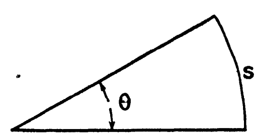


FIG. 7–1

[1] This chapter may be omitted in a short course.

[2] If there were a hundred letters in the English alphabet, physicists would have use for still more to stand as symbols for the variety of quantities with which they deal.

Inertia of rotation. If a gate is to be opened, it must rotate on its hinges. To produce rotation in any body we have already seen (p. 33) that a ***torque*** must be applied to it. If a child is sitting on the gate, it usually does not swing so quickly in response to the same torque, and there is a difference in this respect where the child is, as well as how heavy he is. The inertia of the rotating body depends not only on the mass, or inertia, of its parts, but also on their distances from the axis of rotation. We can see more clearly just what is the nature of this connection if we consider first a very simple case.

Newton's second law for rotation. Rotational inertia. Let a particle m be fastened by a weightless rod of length r (Fig. 7–2) to a center C about which it can rotate; and let a force F be applied to it as shown, tending always to move it in a direction at right angles to the rod. This force will produce an acceleration a, according to Newton's equation $F = ma$, F being measured in absolute units. Since the particle must rotate, the notation of rotation will prove more convenient in describing its motion. Multiply both sides of the equation by r, and we get $Fr = mra$. But Fr is the ***torque,*** which we shall denote by L; and $a = r\alpha$; so that we arrive at the equation $L = mr \times r\alpha = mr^2 \times \alpha$. Call mr^2 by the letter I, and we have $L = I\alpha$, an equation of exactly the same form as the one with which we began. This equation states that the torque which causes the rotation produces an angular acceleration in the particle in a manner entirely similar to the way a force produces an ordinary acceleration in a body of mass m; but, to make this statement true the inertia I must be equal to mr^2. We see from this just how the ***rotational inertia*** of a particle varies with its distance from the axis, namely, as the square of that distance rather than the first power. The quantity I is usually called the ***moment of inertia*** in physics; some engineers, however, apply this term to another quantity of similar form but quite different meaning. On this account, we shall call it ***rotational inertia.*** The units in which it is usually measured are pds.-ft.^{2} in the English system and gm.-cm.^{2} in the metric.

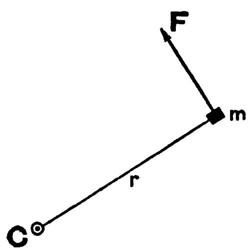


Fig. 7–2

Rotational inertia of hoop and of cylinder. Usually we are not dealing with bodies so simple as a particle. If we have a ***hoop,***

rotating in its own plane about its center, with all its particles at practically the same distance from the center, we may add up the rotational inertias of all the particles, and we shall evidently get Mr^2 for the sum, where M is the mass of the whole hoop. If the body is a solid ***cylinder,*** rotating about its own axis, we have to add together the values of the quantity mr^2 for each of its particles, where the distances r are now variable. This problem is easily done by the methods of calculus, and most students meet with it when they study that subject. The result is $I = \frac{1}{2}MR^2$, where M is the whole mass and R the outside radius of the cylinder.

Radius of gyration. One may express the rotational inertia of this cylinder in another way. Imagine the whole mass to be concentrated at a distance from the center of rotation k, called the ***radius of gyration*** k, such that the rotational inertia remains unchanged. Since $I = \frac{1}{2}MR^2 = M\left(\frac{R}{\sqrt{2}}\right)^2 = Mk^2$, we see that k must be equal in the case of this cylinder to $\frac{R}{\sqrt{2}}\cdot$ A cylinder spinning about any other axis, or any other form of rotating body, requires a different expression for its rotational inertia or radius of gyration. These results will not be derived here, but the most useful ones are listed in the table below, for convenience in solving problems.

TABLE III

Rotational Inertia. Radii of Gyration

	Rotational Inertia (= Moment of Inertia)	Radius of Gyration
Sphere, about an axis through the center	$\frac{2}{5}Mr^2$	$\sqrt{\frac{2}{5}} \times r = 0.632r$
Cylinder, about its own axis	$\frac{1}{2}Mr^2$	$\sqrt{\frac{1}{2}} \times r = 0.707r$
Rod (length L), about axis through middle point, perpendicular to length	$\frac{1}{12}ML^2$	$\sqrt{\frac{1}{12}} \times L = 0.289L$
Rod about axis through end, perpendicular to length	$\frac{1}{3}ML^2$	$\sqrt{\frac{1}{3}} \times L = 0.577L$

Work, energy, and power in rotation. It is easy to obtain expressions for the work done in making a body rotate, and for the kinetic energy which it thus acquires. Consider only the simplest case, a massive particle attached to a light hinged rod, as in Fig. 7–2, page 94, and urged forward by a force always perpendicular to the rod. The ***work done*** is the force multiplied by the arc

through which the particle moves; or Fs. Since $s = r\theta$ (p. 93), the work $= Fr\theta = L\theta$. Note that we should have arrived at this result without proof if we had substituted in the expression Fs for each of these two letters the corresponding term used in rotation. Thus F is what moves bodies in straight motion; L is the torque which makes them rotate; s is the distance; θ is the angle described, and Fs is the same sort of thing as $L\theta$, each being quantities of work.

In a similar way, the particle has a ***kinetic energy*** $\frac{1}{2}mv^2$; but $v = r\omega$, so that the K.E. $= \frac{1}{2}mr^2\omega^2 = \frac{1}{2}I\omega^2$, where again we reach an expression similar in form to the one with which we started, and composed of corresponding terms. This is true in general of rotation; ***to each term and formula in straight motion there corresponds one in rotation, which can be written down by analogy.*** For instance, ***power*** in straight motion is $\dfrac{\text{force} \times \text{distance}}{\text{time}}$, which may be stated as force $\times$ velocity, or Fv; likewise power in rotary motion is $\dfrac{\text{torque} \times \text{angle}}{\text{time}}$, or $L\omega$.

Dimensions of rotation terms. An *angle*, being one length divided by another, has no dimensions. A *torque* is a force multiplied by a distance; its dimensions (p. 76) are therefore ML^2/T^2, which are the same as those of work or energy, though we must remember that the distances involved are in different directions. Multiplying a torque by an angle does not change its dimensions but turns it into work ($L\theta$). An *angular velocity* has the dimensions $1/T$, like a frequency; an *angular acceleration* $1/T^2$. *Rotational inertia* is expressed by ML^2, and the product of this by an angular acceleration yields a torque, as it should according to the equation $L = I\alpha$.

Analogies. It will be a convenience at this point to present a table of quantities and formulæ connected with straight line motion, and a parallel column with the analogous quantities and formulæ connected with rotation. If these analogies are borne in mind, rotation problems can be solved as readily as those already done in the preceding chapters, as examples will show. The similarity in the parallel columns will be noticed throughout, especially in the formulæ applying to uniformly accelerated motion, which we shall consider at once. The table also contains a rotational form of Newton's second law, a kinetic energy expression, and an expression for angular momentum which is not proved but written down by analogy. These will be useful later.

TABLE IV

Straight motion		*Rotation*	
Mass	m	Rotational Inertia	I
Force	F	Torque	L
Distance	s	Angle	θ
Velocity	$v = s/t$	Angular velocity	$\omega = \theta/t$
Acceleration	a	Angular acceleration	α
Work	Fs		$L\theta$
Kinetic energy	$\frac{1}{2}mv^2$		$\frac{1}{2}I\omega^2$
Momentum	mv		$I\omega$
For uniform acceleration			
	$v = at$		$\omega = \alpha t$
	$s = \frac{1}{2}at^2$		$\theta = \frac{1}{2}\alpha t^2$
	$v^2 = 2as$		$\omega^2 = 2\alpha\theta$
	$s = \frac{1}{2}vt$		$\theta = \frac{1}{2}\omega t$
Newton's second law			
	$F = ma$		$L = I\alpha$

Uniformly accelerated rotation. Problems involving uniformly accelerated rotation are to be treated in the same way as those on uniform acceleration in a straight line. As illustrations of this fact, let us consider the following examples.

Example 1. A bicycle wheel, acted on by a steady torque, is observed to make its first revolution from rest in 4 sec. Find the angular acceleration (assumed to be uniform), the time it will take to cover 4 revolutions, and the angular velocity at the end of that time.

A complete revolution is 2π radians; this is θ.

$\theta = \frac{1}{2} \times \alpha t^2$, or $2\pi = \frac{1}{2} \times \alpha 4^2$, whence $\alpha = \frac{\pi}{4}$ radians per second per second.

Four revolutions make 8π radians of angle. Using the same formula again,

$$8\pi = \frac{1}{2} \times \left(\frac{\pi}{4}\right)t^2, \text{ or } t^2 = 64, \text{ and } t = 8 \text{ seconds.}$$

The angular velocity then is

$$\omega = \alpha t = \frac{\pi}{4} \times 8 = 2\pi = 6.28 \text{ radians per second.}$$

Example 2. If a bicycle wheel, moving as in the preceding example, is doing so because it is being pushed by a force of 400 gm. applied in the most effective direction on a spoke 20 cm. out from the center, what must be the rotational inertia of the wheel?

The torque is 400×20 or 8000 gram-cm. (gravitational units) or $980 \times 8000 = 7{,}840{,}000$ dyne-cm. Since the equation $L = I\alpha$ applies in this example,

and is true in this form only for absolute units, it is convenient to reduce to these units.

Since $\alpha = \frac{\pi}{4}$ we have $\quad 7{,}840{,}000 = I \times \frac{\pi}{4}$

whence $\quad I = \dfrac{31{,}360{,}000}{\pi} =$ nearly 10,000,000 gram-cm.^{2}

Example 3. To find the radius of gyration of the bicycle wheel of the preceding example, if its mass is 10 kg.

$$I = Mk^2;\ 10{,}000{,}000 = 10{,}000\ k^2,\ \text{whence } k = 31.6 \text{ cm.}$$

Example 4. Repeating the last two examples, but in English units, with changed data. Force pushing the wheel, 1 lb.; distance of application, 8 in.; mass of wheel, 5 pds.; acceleration as before.

The torque is 1 lb. $\times \frac{2}{3}$ ft. $= \frac{2}{3}$ lb.-ft. Note that the equation $L = I\alpha$ is true as it stands *only* if L is expressed in absolute units (poundals in this case; see p. 65). To convert pounds-force to poundals, we must multiply by a number equal to g or 32.[1] Hence $Lg = I\alpha$ if L is in lbs.-ft., which gives

$$32 \times \tfrac{2}{3} = I \times \frac{\pi}{4} \quad \text{or,} \quad I = \frac{256}{3\pi} = 27.16 \text{ pd.-ft.}^2$$

The radius of gyration is given by $Mk^2 = I$ whence

$$k^2 = \frac{27.16}{5} = 5.432 \quad \text{and} \quad k = 2.33 \text{ ft.}$$

Spin energy. Looking at rotation questions from the point of view of energy often helps us to solve them in a simple manner. As an example consider the following case. Two cylinders (Fig. 7–3) are made of equal size and equal mass, each partly of brass

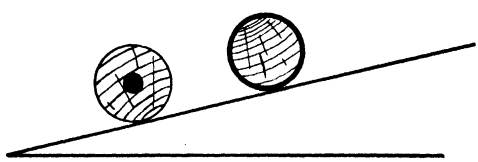


FIG. 7–3

and partly of wood. In one the brass is all near the central axis; in the other all on the outside. Since brass is much denser than wood, its rotational inertia is much greater when it is farther out; as a result the two cylinders possess very different rotational inertias. Now suppose them to roll down a straight slope, starting together from rest. Which one will reach the bottom first?

[1] As used here g is not an acceleration, but a mere number, used to convert one force-unit into another. The "dimensions" (p. 76) of the two sides of the equation would be unlike if g were an acceleration.

At the top of the hill the two cylinders have the same (potential) energy, since they have equal weights and have been lifted to the same height. On the way down this is transformed into kinetic energy of two forms, one due to motion of the center of gravity of the cylinder as a whole, the other due to the spin about its own axis. If they arrived at the bottom at the same time, they would have the same forward speed, and therefore the same rate of spin, since their diameters are the same. But this would mean that the one with the larger rotational inertia would have more spin energy than the other, and therefore more energy altogether, which is impossible. Therefore they cannot arrive at the bottom together, and the one with the brass on the outside must lag behind the other in the race.

Example 1. Consider a case of this sort quantitatively. Let a sphere of any material, of mass M_1, and of radius R_1 be set to run a race against a cylinder of any material, of mass M_2 and of radius R_2, by rolling them down the same incline. As the *sphere* rolls it travels a distance $2\pi R_1$ (its circumference) in one revolution; and in this time it makes a complete rotation, 2π (radians). Its linear velocity v_1 and its angular velocity ω_1 are related by the ratio $\frac{v_1}{\omega_1} = \frac{2\pi R_1}{2\pi}$ or $v_1 = R_1\omega_1$. A similar relation holds for the cylinder. The sphere has had stored in it, by being lifted to the top of the hill through a vertical height of h cm., a quantity of potential energy equal to M_1gh ergs (if C. G. S. units are used). In the journey down the hill this has been transformed into kinetic energy in two forms, forward motion and spin. Hence, by the principle of the conservation of energy

$$M_1gh = \tfrac{1}{2} \times M_1v_1^2 + \tfrac{1}{2} \times I_1\omega_1^2.$$

But, for a sphere $\quad I_1 = \tfrac{2}{5} \times M_1R_1^2. \quad$ (Table III)

Hence $\quad M_1gh = \tfrac{1}{2} \times M_1v_1^2 + \tfrac{1}{2} \times \tfrac{2}{5} \times M_1R_1^2\omega_1^2.$

Also, substituting v_1 for $R_1\omega_1$, we have

$$gh = \tfrac{1}{2} \times v_1^2 + \tfrac{1}{5} \times v_1^2 = \tfrac{7}{10} \times v_1^2$$

or $\quad v_1^2 = \tfrac{10}{7} \times gh = 1.429gh.$

We see at once that the mass and the radius have canceled out; in other words all spheres roll down at the same rate.

For the *cylinder*, $\quad I_2 = \tfrac{1}{2} \times M_2R_2^2 \quad$ (Table III)

and $\quad gh = \tfrac{1}{2} \times v_2^2 + \tfrac{1}{4} \times v_2^2 = \tfrac{3}{4} \times v_2^2$

or $\quad v_2^2 = \tfrac{4}{3} \times gh = 1.333gh.$

Hence $\quad v_1 = 1.036v_2.$

All cylinders roll down the hill at the same rate, irrespective of their size, but spheres go faster than cylinders, and we see just how much faster. If the plane

The whirling-table experiment. A curious experiment depends on the last statement for its explanation. A round iron platform is very carefully mounted on ball bearings so as to turn with almost no friction. A man stands on it with his arms outstretched, holding heavy weights in his hands (Fig. 7–4). If the platform is properly built and leveled, he can do nothing to start his whole body in continuous rotation. If he turns his head to the right, his body will turn to the left, so as to keep the resultant angular momentum always zero. If he waves one of the weights in a circular fashion over his head, his body will rotate in the opposite sense so as to maintain a zero resultant momentum, as before. Now let another man apply a small torque to him. He will then begin to rotate steadily at a moderate rate. When the external torque has ceased, it will have stored in him a definite amount of angular momentum, which he will retain unchanged so long as no more torque is allowed to act on him. Write this amount as $I\omega$, implying by the sizes of the letters that his rotational inertia is large, on account of his extended hands and the weights, but his angular momentum small. Now let him bring his hands in toward his neck, as close as he can. As he does so he will be diminishing his I, but his $I\omega$ will remain constant. The result we can write as $\iota\Omega$, implying that his rate of spin will greatly increase. On putting his arms out again he will slow down to the original rate of spin once more. No forces that he can exert on himself can alter the total angular momentum which he possesses, but he can vary his angular velocity, by changing his moment of inertia.

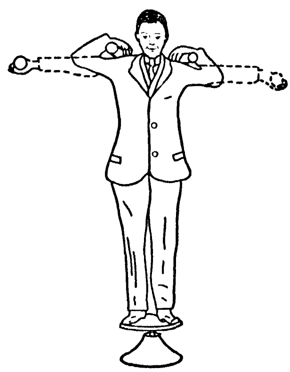


FIG. 7–4

Another way of explaining this experiment simply is to say that when the weights are well out they have a certain linear momentum, which they tend to retain as they are brought in, in accordance with Newton's laws. But, on coming closer to the axis of of rotation they reach a place where the linear speed of that part of the man's body is less. Therefore the weights will give some of their momentum to the man's body and speed it up a little; so that the incoming weights will pull him into a more rapid spin.

Energy changes in whirling-table experiment. It is interesting to consider the energy changes in this case. It is evident that if $I\omega$ remains constant and ω changes, $I\omega^2$ must also change. The kinetic energy therefore increases as the weights come in. It is easily seen that the man does work on pulling the weights in against centrifugal force, and this is the source of the extra kinetic energy.

Another form of the experiment. An interesting variation of the whirling-table experiment is furnished by giving the man a bicycle wheel to hold whose axle has been extended to form a handle. If a band of lead is wound around the rim in place of the usual tire, a body of large rotational inertia is formed. If the man stands at rest on the whirling table, holding the axle of the bicycle wheel horizontal, he may set it spinning as fast as he pleases without affecting his own condition, because he cannot spin around a horizontal axis. But if, when it is spinning, he turns the axle of the wheel so that it is vertical, or partly so, the wheel will thereby acquire some angular momentum around a vertical axis, and the man must at the same time acquire an exactly opposite amount. He may thus make himself spin in either direction by merely tilting the wheel in his hand, and yet at all times the system composed of him and the wheel together possesses zero angular momentum about a vertical axis.

The rotating earth. The earth spinning on its axis presents a problem not unlike that of the man on the whirling table. If it should shrink, as seems not unlikely if it is cooling off internally, its rotational inertia will diminish, and its rate of revolution must increase on this account. Of course, in reality other causes also may alter it; the tides, for example, in rushing over the sea bottom create a force of friction which tends to diminish the rate of revolution and complicate the problem. Experimental evidence is now being obtained in various ways which indicates that the earth's rate of rotation is not perfectly constant.

Foucault's pendulum. The rotation of the earth can be demonstrated by means of a device first used in 1851 by Foucault.[1] He hung a long pendulum with a heavy bob from the top of the dome of the Pantheon in Paris, and set it swinging in a direction which was carefully marked. In a short time it was plainly seen to have changed its direction, and the plane of its motion continued to turn at the rate of about 11° per hour. It is easy to see that the pendulum, if it is regarded as a body possessing angular momentum, cannot change the plane of its motion unless a torque acts on it; but there is no way in which such a torque can be supplied if the pendulum is properly mounted. Hence the plane of the motion must remain unchanged, in spite of rotation of the earth. If the experiment can be imagined to be carried out at the north pole, the earth will rotate at the rate of 15° per hour below the pendulum, the

[1] J. B. L. Foucault, 1819–1868, French physicist, who investigated many interesting optical problems, discovered "eddy" (or Foucault) currents induced in a piece of metal near an electromagnet and measured the velocity of light in various media.

latter's plane remaining fixed in space. At the equator there will be no such turning; while at the south pole it must occur oppositely. At intermediate latitudes the rotation can be shown to be reduced in proportion to the sine of the angle of latitude. This behavior of a pendulum can be shown in the laboratory provided one furnishes a rigid support which imposes no tendency on the pendulum to vibrate more easily in one direction than in any other, and the wire should be very straight, which is best done by making the load as heavy as it can safely support.

The gyroscope. The spinning top presents another curious problem. It seems to a child like magic to see that a top does not

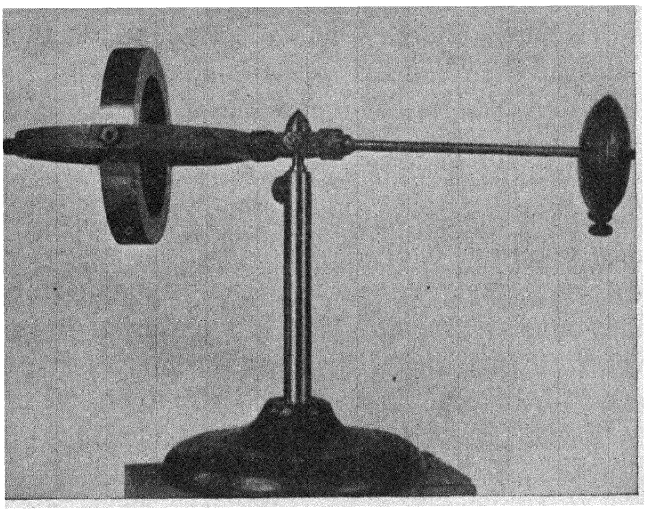


FIG. 7–5
A gyroscope

fall over when spinning in a slanting position; and yet the explanation is not difficult. We shall consider the similar case of the gyroscope. A heavy disc with good bearings is carefully balanced over a pointed stand by an adjustable counterpoise, as shown in Fig. 7–5. If the disc is now set spinning nothing unusual happens, but if the counterpoising weight is set a little farther in or out, the apparatus begins to revolve slowly ***in a horizontal plane.*** This motion is called ***precession.*** To explain it we must go back to a combination of two vectors.

Angular velocity as a vector quantity. An angular velocity has both amount and direction, the two features of a line. It is a vector quantity, and can be handled in the same way as a force, or any other vector. The only novel feature it presents is in the

choice of the direction in which the line representing it is to be drawn. There is evidently no line characteristic of a spin except the line of its axis. We make an arbitrary agreement as to the direction in which this line should be drawn, by remembering that on turning a screw driver in the direction of the motion of the hands of a clock ("clockwise") we drive the point of the screw forward. So we draw an angular velocity as a line facing ***away*** from us if it is to represent a spin in the ***clockwise*** direction. Finally we make the line of a suitable length to represent the amount of the angular velocity measured in radians per second, on some convenient scale; Fig. 7–6 illustrates the vector line of a spin.

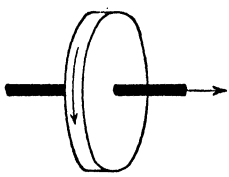


Fig. 7–6

Angular momentum may equally well be considered a vector quantity and represented by a properly chosen line.

Combination of angular velocities. We have already found it a convenience to be able to treat vectors by the parallelogram rule. We can do this for angular velocities also. Imagine, for instance, how a body must be moving if it is spinning at a certain rate around a vertical axis, and at the same time at a rate twice as fast around an east-west axis. It is not at all easy to picture in the mind a combination of two such motions. But, if we draw the corresponding vectors, we get a rectangle with one side twice as long as the other, and the diagonal gives us the new combined direction of spin, and its rate. Thus the problem is simplified and can be solved by familiar methods. These ideas enter into the explanation of the precession of the gyroscope to which we now return.

The precession of the gyroscope. Suppose that the gyroscope disc, seen from the position of the pivot, is spinning in a counterclockwise direction. Then the vector representing its angular momentum (which is better to use here than velocity) will run to the right along the rod which supports the disc, and will be a long one, since the rate of spin is supposed to be quite rapid. Now let us look at the gyroscope from the side, as in Fig. 7–5, and let it be unequally counterpoised, the disc end being too light. If originally horizontal, it will now begin to tip upwards. This tipping will generate a small angular velocity around an axis perpendicular to the paper, and inward, since the sense of the rotation is clockwise, as seen by us. Now, looking down from above on the gyroscope, let us

draw the two vectors representing its two spins. AB (Fig. 7–7) is the vector representing the angular momentum of the disc, and AC is the new vector representing the angular momentum due to the tipping of the apparatus. These will automatically and instantly combine into AD, and this combination will occur long before AC gets to be large enough to be seen, and will keep on occurring continually. The result will be that the axis of the gyroscope, instead of tipping down, will remain almost horizontal, but will slowly turn in the direction of AD, that is (still looking down from above on the apparatus) the gyroscope will precess in a counterclockwise direction. If the counterpoise is now slid in to a point where it tends to make the spinning disc fall, the direction of AC will be reversed, and the direction of the precession will reverse with it.

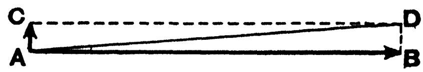


Fig. 7–7

The gyroscope when held in the hand feels like a live thing. If we try to make it turn one way it turns at right angles to what we expect. If when it is mounted on its pivot, spinning as before, we try to accelerate its precession by pushing on it with a vertical pencil, it rises instead. If we try to make it rise, its precession slows down.

Precession occurs also in airplanes and automobiles. If an ordinary automobile is turning to the right, with its engine spinning counter-clockwise as seen from the driver's seat, the precession tends to lift the front of the car and depress the rear, as can readily be verified by drawing the vector rectangle.

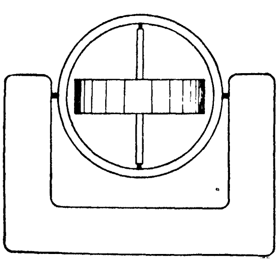


Fig. 7–8
A gyroscopic ship-stabilizer

The gyroscope has been cleverly applied to ***stabilizing a ship*** in a rough sea. A heavy gyro-wheel is mounted on a vertical axis which on its own part is free to turn about a horizontal axis transverse to the ship, as shown in Fig. 7–8. When the ship starts to roll, the torque tending to tip the axis of the wheel along with the rest of the ship produces precession, the axis leaning over toward the bow or the stern, depending on the direction of spin and of rolling. A large torque may be needed to do this, and therefore, as a reaction to this, the gyro-wheel is able to exert a large

torque on the ship, opposing the rolling. If friction is introduced to check the precession, the rolling of the ship will thereby be reduced. A ship that might otherwise roll through an angle of 30° can thus be held steady to within less than 4°. Figure 7-9 shows a model of such a device due to Franklin,[1] which has a wheel mounted on a board, cut out to resemble a ship. The board has only two feet to rest on but when the wheel is spinning it stands up on them and strongly resists a rolling motion. A large force tending to tip the model merely produces a rapid precession of the gyroscope about its horizontal axis, while the model remains level.

The ***gyro-compass*** is another useful invention; it is free from the defects of the magnetic compass, (p. 300) which are particularly

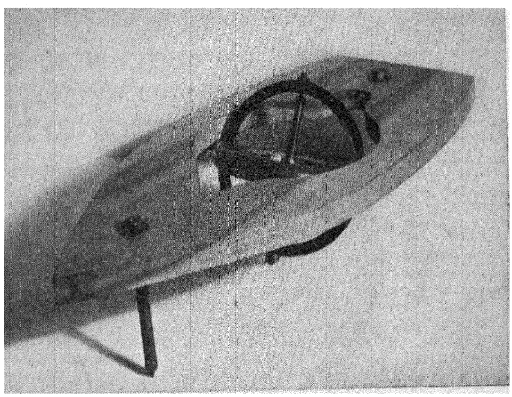


FIG. 7-9
A model of a gyroscopic ship-stabilizer

troublesome on war vessels on account of the amount of steel they carry. It has, however, some peculiarities of its own, such as being slightly disturbed when the ship has a northward or southward velocity. Its wheel is kept spinning by an electric (induction) motor, and is mounted on an axis which is constrained to lie in a horizontal plane by being free to swing, somewhat as a pendulum does. If this axis is not pointing northwards, the rotation of the earth tips it and this tipping produces precession, the axis turning slowly into a north-south line. This may take two or three hours when the wheel is first started, but thereafter it maintains its proper direction quite closely. It is to be noted that this direction

[1] W. S. Franklin, late professor in Mass. Institute of Technology; author of many original and interesting books on physics and electrical engineering.

is the true geographic north, rather than the magnetic, which is one of the greatest advantages of this type of compass.

Precession of other bodies. The tendency of a spinning body to keep the axis of spin unchanged is analogous to the tendency of bodies moving in straight lines to keep on, as in Newton's first law. It takes a torque to change the direction of the axis of spin. A rifle bullet is given a spin, so that its nose will remain pointing forward. A heavy shell from a battleship's gun starts in the same manner, but as it follows the curve of a projectile, its nose is soon pointing too high. Air resistance then begins to apply a torque to it, under which it moves at right angles to the way one might otherwise expect, in a sort of precessional motion, wobbling as it goes. The earth also in its passage around the sun moves with a precessional motion. As it is not perfectly round, the sun's attraction exerts a torque upon it in addition to the gravitational force pulling at its center. The precession is very slow and has the effect of making the north pole, the point in the heavens toward which the axis of the earth is directed, travel about in a circle of large diameter, making a complete revolution in about 26,000 years.

Reference books:

H. Crabtree. "Spinning Tops and Gyroscopic Motion," 1914 (Longmans, Green).

J. Perry. "Spinning Tops," new edition, 1929 (Sheldon Press)

A. M. Worthington. "Dynamics of Rotation," 1925 (Longmans, Green).

PROBLEMS

1. Calculate the angular velocities of the hour, minute, and second hands of a watch.

2. Explain why it is possible to pull a spool along a table by the end of the thread which is wound on it so that (*a*) the spool spins forward; (*b*) drags without spinning; (*c*) spins backward.

3. Make a vector diagram showing how precession occurs in the case of the gyroscopic ship stabilizer.

4. If an airplane, whose propeller is turning clockwise as seen from the front, is tipped down in order to make a descent, what precessional effect will occur, and how will this affect the steering? How would this be altered if there were two propellers turning in opposite directions?

5. A uniform circular disc, 10 kgm. mass, 10 cm. radius, has a light shaft of 1 cm. diameter projecting along its axis in both directions. It rolls on this

shaft down an inclined run-way formed of two parallel rods. If the disc descends 10 cm. vertical height by rolling down on the rods, find (*a*) the loss of potential energy in ergs, and (*b*) the linear velocity of the center of gravity of the disc at the end of this motion.

6. A disc of 24 pds. mass and radius of gyration 1 ft. is pivoted around an axis through its center. It has an axle of 2 in. diameter around which a thin string is wrapped. A force of 2 lbs. is applied to the free end of the string, starting the disc into rotation. Find the torque due to this force, the angular acceleration, and the time of the first revolution.

7. The front wheel of an automobile is jacked up, and is free to rotate without appreciable friction. The mass of the wheel is 40 pds., and its radius of gyration is 1.5 ft. If one of its spokes is pushed at right angles to itself at a point 8 in. out from the axis with a force of 3 lbs., the wheel starts from rest with a constant angular acceleration. Find the angular velocity acquired in 5 sec., and the number of revolutions in this time.

8. A spherical ball of 5 cm. radius rolls down an inclined plane whose height is 10 cm. and length (along the incline) 100 cm. By equating the energy at the top and at the bottom, find the linear speed of the ball at the bottom of the incline.

9. A 5 H.P. engine is connected to a flywheel whose whole mass (2200 pds.) may be regarded as concentrated in its rim at a distance of 4 ft. from the axis. Assuming no friction, find the time required by the engine working at its full power to bring the wheel from rest into rotation at a speed of 1 revolution per second.

10. A grindstone, mass 256 pds., radius 1 ft., turns on a frictionless bearing at the rate of 2 revolutions a second. A tool pressed against its circumference exerts a retarding frictional force of 2 lbs. Find the time before it comes to rest.

11. A bicycle wheel, mass 5 kgm., (assumed to be all in its rim at a radius of 30 cm.), is set spinning at a rate of 2 revolutions a second while it is held by its axle. It is then set down upon an incline, so as to run up it. If the incline rises 1 ft. vertically for each 10 ft. along the slant, and if none of the spin energy is lost in other ways, how far up the incline will the wheel run?

CHAPTER 8

GRAVITATION

Newton's law, 109; the Cavendish experiment, 109; another method, 111; properties of gravitation, 111; the mass of the earth, 111; variations of gravity, 112; gravitational units of force, 113; other deductions from Newton's law of gravitation, 114; the tides, 114; tides in the earth and air, 115; tidal power, 116.

Newton's law. It is one of our earliest experiences that things fall to the ground if we let them go; and yet the reason for this action remains a mystery, the most familiar mystery in physical science. Newton (1687) made a great advance by finding that falling is due to a force which is of a most general sort, by means of which every body in the universe appears to be able to ~~attract~~ every other body. ***Newton's law of gravitation*** is expressed by

$$F = G\frac{mm'}{r^2},$$

where F is the attracting force between the two bodies whose masses are m and m', and whose centers of gravity are a distance r apart. G is a very small quantity, called the constant of gravitation. It is numerically equal to the minute force of attraction between two one-gram masses whose centers are 1 cm. apart, as is easily seen by putting $m = m' = 1$ and $r = 1$ in the equation. Its ~~numerical value depends~~ on the units used; it is 6.664×10^{-8}, if dynes are used, and 6.80×10^{-11} if grams of force are used (assuming $g = 980$). G is not a pure number, but is an example of a peculiar class of quantities; it is a constant which has dimensions, so that its value changes with the units used.

The Cavendish experiment. The measurement of the value of the constant of gravitation is a very delicate matter, first carried through successfully by Cavendish [1] (1798), later much improved

[1] Henry Cavendish (1731–1810), a man of wealth and noble family who preferred a life of almost fantastic isolation, and occupied himself with chemical and physical ideas. He first discovered the nature of hydrogen and of water, and recognized the existence of what we now call specific and latent heat, though he never obtained recognition for many of his ideas, as he preferred to leave them unpublished.

by Boys[1] (1895) and recently carried to a still higher degree of precision by Heyl.[2] In one form of this experiment the arrangement is as follows: — a very fine but strong fiber made of fused quartz, *FG*, supports a light horizontal rod (Figs. 8–1 and 8–2) from the ends of which are hung small lead or gold balls by means of wires of unequal length. Large lead balls *M* and *M* are supported near the small masses *m*, in such positions that the attractions of each *M* and its corresponding *m* tend to twist the fiber in the same sense. Any turning of the rod is detected and measured by a beam of light thrown upon a small mirror *G* fastened to the rod, which reflects light to a scale, or into a telescope. The position of the reflected light is observed when the masses are placed as shown in the figure. Then the large masses are moved so that each one is on the opposite side of its small mass, and the same distance away. The large masses in either position turn the rod through a small angle from its neutral position; reversing the large masses makes the rod turn through twice this angle. Having masses at each end of the rod produces twice as much torque on the rod as a single pair of masses would. Thus this arrangement and manipulation of the masses greatly increases the very small

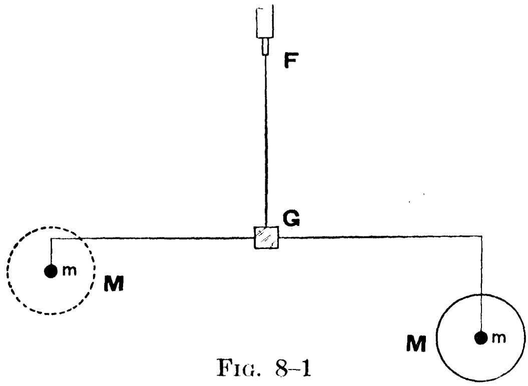


FIG. 8–1
The Cavendish experiment, viewed from the side

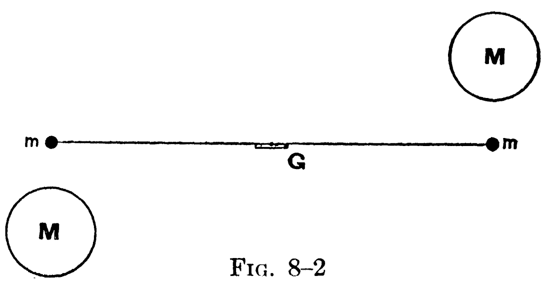


FIG. 8–2
The Cavendish experiment, viewed from above

[1] Professor C. V. Boys, English physicist, to whose skill we owe the method of drawing fine quartz fibers which have been so useful in delicate physical instruments; author of an interesting little book on "Soap Bubbles," 1890 (Society for Promoting Christian Knowledge).

[2] Dr. Paul R. Heyl, of the Bureau of Standards, Washington, D. C.

effect produced by the gravitational forces. A knowledge of the torsional stiffness of the quartz fiber can be obtained from its dimensions and coefficient of elasticity (see p. 123). From this, combined with the angle of turning of the rod, the actual amount of the force of attraction F between one mass M and the little mass m can be found. The distance r between the centers of the balls is easily measured. Hence Newton's formula can be tested, and the value of the constant G obtained. This arrangement of the apparatus avoids almost entirely any action of one large mass M upon the small mass m at the far end of the rod; this is the reason for supporting the little masses at unequal heights.

Another method. It is also possible to measure the constant of gravitation by means of a very delicate balance of the type used by chemists. If one places equal weights, as large as the balance will properly bear, in the pans and then places a very large lead weight immediately under first one pan, then the other, the added attraction produced by the lead weight on the weights in the pans can be measured, and from the dimensions of the experiment the constant of gravitation can be calculated.

Properties of gravitation. Newton was able to prove that the moon is held in its path around the earth, in accordance with his law, by the force of gravitation, which just balances the centrifugal force. Astronomers use the law of gravitation constantly in working out the paths of the heavenly bodies, and it evidently applies also to bodies in our laboratories. We have good reason to suppose it to hold universally. The action of gravitation takes place freely across spaces which we believe to be quite empty of ordinary matter. It makes itself felt through other matter, and it appears to be practically instantaneous in its action between distant bodies. The suggestion that it may be propagated with the same velocity as light and electrical effects is attractive and may be correct. As yet no method of measuring its speed has been found. When we are dealing with the attraction of the earth upon bodies at or near its surface, we call the action ***gravity,*** reserving the term ***gravitation*** for more general use.

The mass of the earth. If one attracting mass is the earth itself, and the other is a 1-gm. mass in the laboratory, Newton's law states that

$$980 \text{ dynes} = 6.664 \times 10^{-8} \times \frac{\text{mass of earth} \times 1 \text{ gram}}{(\text{radius of earth in cm.})^2}$$

since the pull of the earth on a gram mass is 980 dynes. If the radius of the earth is 6.37×10^8 cm., the mass of the earth can be found directly from the formula. It comes out about 6.0×10^{27} grams, and since its volume is 1.08×10^{27} cubic centimeters, the average density is found to be 5.515 grams per cubic centimeter or about twice as much as the average density of the material in the earth's crust that we can reach. This goes well with the suggestion obtained from other sources that the central portion of the earth contains a great deal of iron.

Variations of gravity. Bodies at the surface of the earth, or above it, are attracted to it as though all its mass were concentrated at the center. A proof of this fact was given by Newton. The diminution of the force of gravity with increase of distance upward from the earth's surface is, therefore, not very rapid, as bodies at the surface are already so far from the center. Since the earth bulges out some 13 miles at the equator, instead of being perfectly round, the force of gravity there is less than at the poles, though centrifugal force also affects its value. Actually, the force of gravity at the poles is about 0.5% greater than at the equator, and centrifugal force is responsible for about 0.3%, leaving 0.2% to be accounted for by the shape of the earth. If we go down in a mine with the idea of getting a little nearer to the center, we find the force of gravity diminishing instead of increasing. We can see that it must do so from the fact that at the center of the earth a body would be pulled equally in all directions, and would in other words have no weight at all.

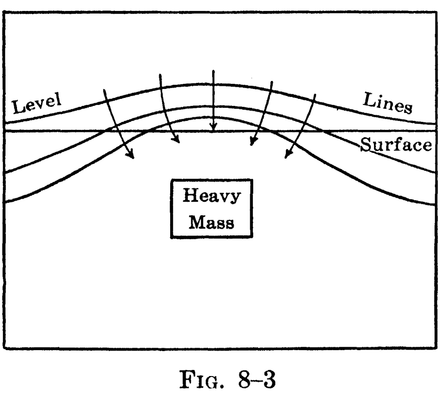


Fig. 8-3

A mountain rising abruptly out of a flat plain has enough sidewise gravitational action to make a plumb bob on the plain hang perceptibly out of the vertical, leaning over toward the attracting mass through a small but measurable angle. Minute variations in the force of gravity both in direction and amount are now detectable by refined methods, including one using an extremely delicate "torsion balance," invented by Eötvös, which is capable of measuring horizontal forces whose amounts are to the force of

gravity as one in a million million (10^{12}). These methods are being extensively applied at the present time to the discovery of oil and other valuable deposits. Figure 8–3 illustrates in exaggerated form the changes in the force of gravity due to a heavy mass, discovered by such devices.

A few values of the acceleration of gravity are collected in Table V, showing the extent of the variation in different places. These variations are due not only to lack of perfect roundness of the earth, or to centrifugal force, but also to differences in height above sea level, to variations in density of the crust of the earth, and perhaps to many other causes. The usual way of obtaining these results is discussed in Chapter 16.

TABLE V

Acceleration of Gravity

Minneapolis, Minn.	980.60
Boston, Mass.	980.39
Washington, D.C.	980.10
San Francisco, Cal.	979.97
New Orleans, La.	979.32
Key West, Fla.	978.97
Arctic Sea, Lat. 85°	983.17
Montreal, Canada	980.65
Greenwich, England	981.19
Rome, Italy	980.35
Cape Town, Africa	979.66

Gravitational units of force. The pound-force and gram-force are variable quantities to the extent indicated in the preceding paragraph. As exact scientific units of measurement they would be unsatisfactory unless fixed values were adopted for them. The "standard" weight is taken to be the value at sea level at 45° latitude, where the average value of the acceleration g is taken to be[1] g = 980.665 cm./sec.^{2} or 32.1740 ft./sec.^{2} The standard pound weight in England is the weight of 1 pound mass at London. The weight at any other place can be found from the value of g at that place. Fortunately, in ordinary weighing, we make ***comparisons,*** by beam balances, whose results are the same in any place, since equal masses have equal weights anywhere. But, if we use spring balances, we get results which vary as we move about. It

[1] Recent work indicates that 980.616 is nearer the truth.

is observed, however, that the variations are small, and anyone content with the inaccurate results usually given by spring balances will not consider them important.

Other deductions from Newton's law of gravitation. The mass of the earth has already been found, but it can also be determined in another interesting way. If we assume that the earth's mass is M, that of the moon m, the moon's distance R, and further suppose that the moon is moving uniformly around the earth in a circular path (which is approximately true), the force of gravitation holding the moon in its orbit must be equal to the centrifugal force, or

$$G\frac{Mm}{R^2} = \frac{mv^2}{R}$$

where v is the linear velocity of the moon. But v is the circumference of the orbit divided by the time T which the moon takes for one revolution, or

$$v = 2\pi\frac{R}{T}.$$

Hence,

$$G\frac{Mm}{R^2} = \frac{m \times 4\pi^2R^2}{RT^2} \quad \text{or} \quad \frac{GM}{4\pi^2} = \frac{R^3}{T^2}.$$

This result means that if we measure R and T, and know G from laboratory measurements, we can obtain M, the mass of the earth. This can also be done for any other planet, or star, which has a satellite revolving about it, provided the distance between them and the time of revolution can be found. For instance, the equation gives us the mass of the sun if we make R and T refer to the earth, revolving about the sun. If we have several bodies revolving about a central one, as the planets about the sun, the equation says that the right-hand side is constant, that is, R^3 is proportional to T^2; or, the cubes of the radii are proportional to the squares of the periodic times. This is one of *Kepler's* [1] laws, (1618) which helped to establish the fact that the earth was not the center around which the sun revolved, and hence did much toward turning the attention of mankind toward experimental science.

The tides. A noteworthy example of gravitational effects is furnished by the tides. These are most conspicuous in the oceans, though they can be observed in lakes, and even in the solid earth. The cause of tides is best understood by imagining a simplified case. If the earth were not rotating and were covered uniformly with a deep ocean, the portion of that ocean nearest to the moon (A, Fig. 8–4) would be attracted by that body with a larger force than a corresponding quantity of matter situated at the center of the earth; and this in turn would be more attracted than the

[1] Johann Kepler, (1531–1630); German astrologer and astronomer, whose active and turbulent life led him to many places and positions. His ideas were a curious mixture of mysticism, metaphysics and exact science.

ocean still farther away at *B*. The differences of these forces of attraction furnish tide-raising forces which are very small, of the order of one ten-millionth of the weight of the water. The sun's attraction creates a similar tide-raising force roughly half as great. The moon's force raises a tide which repeats itself twice in one day and 51 minutes. The solar tides, of course, repeat themselves twice in one day exactly. The combination of the two makes a rather complicated motion, which is made more so at any one place by a variety of other circumstances, such as irregularities in depth and shore line, and changes in the height of the moon and sun as seen from the place in question. At times the sun, moon, and earth are in line, and then the effects are greatest ("spring" tides). The tides may at these times range through a height twice as great as at others.

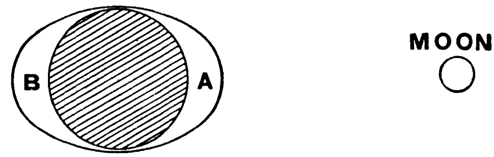


FIG. 8–4
Tides on non-rotating earth

The picture given in Fig. 8–4 is misleading when applied to a rotating earth. The tide-raising force upon the oceans is greatest when the water is under the moon, but the water continues to rise after the time when the moon is at its greatest height above the place of observation. This is in part due to the inertia of the water, an effect which should, according to theory, produce a 90° (or six-hour) lag of the high water as in Fig. 8–5 behind the time when the moon is highest. On account of friction and other causes the actual delay may be much more.

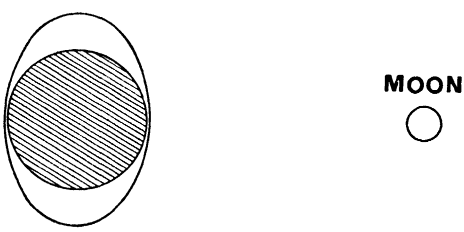


FIG. 8–5
Tides on rotating earth

Tides in the earth and air. The above reasoning would lead one to expect tides in the earth also, if it is capable of yielding at all to tidal forces, which of course it is, not being infinitely rigid. The actual rise and fall of the earth's surface has been measured at Chicago by Michelson and Gale, and found to be as much as nine inches at the time of the spring tides. The amount of this effect indicates that the body of the earth is about as rigid as it would be if made of steel.

The tide-raising force alters the atmospheric pressure by producing tides in the earth's atmosphere. The effect is very small, a thirtieth of a millimeter change in the barometric height, or less.

Tidal power. Attempts to derive power from the tidal flow have been successfully made on a small scale, but the supply is intermittent and the hours when it is available are often inconvenient.

A large development of tidal power would be possible in a region where islands could be connected by dams, so as to produce two large reservoirs. The water could pass through wide openings into the "upper" reservoir at times of high tide only. This could then flow through the power plant into the other reservoir, which must be emptied rapidly at low tide only. The size of both reservoirs must be sufficient to ensure a continuous flow through the power plant. Such a project has been planned for the region where Maine and New Brunswick meet the sea, and where the tides are high. It is unfortunate that favorable natural conditions are not common near the large centers of population.

Simple calculations show that the tides are now wasting power in friction at the rate of 2,100,000,000 horse power, a good deal of it, however, in places inconveniently far north. This friction, combined with other causes, is slowing down the revolution of the earth so that the length of the day is increasing by about one-thousandth of a second in a century.

PROBLEMS

1. Find the change in the weight of a 1-lb. mass if it rises in a balloon to a height of 4 mi. above the surface of the earth. (Radius of earth = 4000 mi.)

2. In one form of the Cavendish experiment the large mass of lead had a mass of 5000 grams, the small ball of 10 grams, and the distance between their centers was 7 cm. The force of attraction between them was found to be one sixteen-millionth of a gram of force. What value does this yield for the constant of gravitation?

3. A mountain rises out of a plain to a height of 1500 meters. It is equivalent in its gravitating action to a cubical mass 1000 m. on a side. Assuming it to be made of material whose average density is 2.5, find the gravitational force it would exert on a 1-gram mass situated on the same level as its center of gravity and 1000 m. away.

4. Calculate the gravitational force between the earth and the moon, given the following data: mass of earth 6.0×10^{27} grams; mass of moon 7.3×10^{25} grams; distance between their centers 3.84×10^{10} cm.

5. How much would a man of 70 kg. mass "weigh" on the surface of the moon, if the mass of the moon is 7.3×10^{25} grams and its radius 1.7×10^{8} cm.?

6. Should the value of gravity on the earth depend on the position of the moon and sun? Calculate the attraction between a 1-gram mass on the earth's surface and the moon, assumed to be straight overhead, using the data given in the preceding problems.

7. Calculate the mass of the earth in grams, given $G = 6.6 \times 10^{-8}$ (in dynes), and the radius of the earth 6.4×10^{8} cm.

CHAPTER 9

ELASTICITY AND SURFACE TENSION

Boyle's law, 118; McLeod gauge, 119; elasticity, 119; elasticity of gases, 120; elasticity of liquids, 120; elasticity of solids, 121; space lattices, 121; Hooke's law, 122; fundamental elastic deformations, 123; other elastic coefficients, 123; spring and torsion balances, 124; the range of elastic forces, 125; the crystalline structure of metals, 125; elastic lag, 126. Surface tension, cause of surface action, 127; surface tension, 128; the spreading of oil over a surface, 130; changes in surface tension, 131; Rayleigh's determination of the size of an oil molecule, 131; adhesion and cohesion, 132; rise of liquids in tubes, 133; depression in tubes, 134; pressure in drops, 135; tension inside a liquid, 135; pseudo-biological effects, 136.

Boyle's law. Anyone who has enjoyed the benefits of an air-filled cushion, or automobile tire, owes these to the property of air, shared in common with all gases, which is called its elasticity. The more the gas is compressed the more it resists compression. This fact is known as Boyle's law[1] (1660). Stated more formally this law says that the pressure p in a gas varies inversely as the volume v, or the product pv is constant. This is true only if the gas is kept at the same temperature while being measured (p. 160), and if the range of pressure over which the test is made is a moderate one.

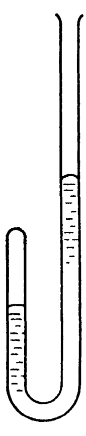


FIG. 9–1 A Boyle's law tube

Tests of this law form the object of a common laboratory experiment. In its simplest form a quantity of air (or other gas) is trapped by a column of mercury in the closed side of a long U-shaped tube (Fig. 9–1), open only on one leg of the U. Addition of mercury to the open end increases the pressure on the gas, and diminishes its volume. The volume is proportional to the length of tubing occupied by the gas. The pressure is equal to that of the atmosphere ***plus*** that due to the excess height of mercury in the open tube over that in the

[1] Robert Boyle (1627–1691), English natural philosopher and alchemist. He discovered the action of the air in transmitting sound, and made many experiments in other parts of physics.

closed. Measurements with an ordinary meter bar are just accurate enough to disclose the fact that Boyle's law represents the behavior of many gases only approximately; at low pressures, however, the agreement is exact.

McLeod gauge. An interesting application of Boyle's law is furnished by the McLeod gauge for the measurement of low gas pressures. The ordinary mercury manometer (p. 6) is much used to measure pressures that are moderately low, but when the difference in height of the two mercury surfaces is as little as a millimeter, for instance, it cannot be read with much precision, and magnifying it by microscopes is inconvenient. In the McLeod gauge, the pressure is magnified before it is measured. The vessel in which the partial vacuum exists is connected by the tube V (Fig. 9–2) to the chamber C, which ends above in a narrow, closed graduated tube T. A reservoir of mercury R connected to the lower part of the apparatus by a flexible tube can be raised, and thus mercury is made to rise past the junction J and cut off in C a sample, so to speak, of the pressure to be measured. The reservoir is then raised until the left-hand mercury surface is at some mark in T. The gas trapped there is under a much higher pressure than before, and the right-hand mercury level, which is now in tube V, serves to measure it. This pressure will be greater than the original one in the inverse ratio of the volumes occupied by the trapped gas (C and T), which are not difficult to measure once for all at the time when the apparatus is made.

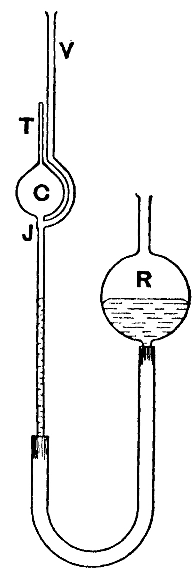


FIG. 9–2
A McLeod gauge

Elasticity. Elasticity is a term used in various senses. In ordinary speech a "very elastic substance" (e.g. a rubber band) is one that yields a great deal to distorting forces, and yet returns to its original form afterwards. In technical language, on the contrary, a "substance with high elasticity" (e.g. a bar of steel) is one which resists distorting forces vigorously. A "perfectly elastic substance" is one that responds alike to compression and to expansion, and returns perfectly to its original form when the distorting force is removed. A watch spring might serve as an example.

In comparing the elasticity of different substances, and in calculations, it is convenient to use "coefficients" of elasticity. A coefficient is defined as the ***stress*** divided by the corresponding ***strain***. In general terms, the ***stress*** is the force per unit dimensions, and the ***strain*** is the effect produced per unit dimensions; for ex-

ample, if the effect is a compression, the stress is the compressing force per unit area (or the pressure), and the strain is the shrinkage per unit volume; their ratio is the volume ***coefficient of compression*** which is one sort of coefficient of elasticity. For most substances the changes are usually quite small, and within this narrow range the behavior of a given substance is regular.

Elasticity of gases. The only deformation that gases resist is a compression; their only coefficient is the compression (sometimes called volume) coefficient. If the compression is slow, the temperature remains constant and Boyle's law is followed. If P is the pressure, p the small increase of pressure producing the change v in the volume V, then the change in volume per unit volume is v/V, and the compression coefficient is $\frac{p}{v/V} = \frac{pV}{v}\cdot$ But, by Boyle's law, $PV = (P + p)(V - v)$, from which it follows that $pV = Pv$, if we neglect the very small product pv. Hence $pV/v = P$, or, the ***coefficient of compression of a gas is equal to the pressure itself.*** If a gas is compressed quickly, so that its temperature rises (p. 219), the pressure increases faster. The coefficient may then be increased by a factor γ (gamma), which is the ratio of the two specific heats of the gas (see p. 170) and has a value for common gases near 1.4.

Elasticity of liquids. Liquids, like gases, have no shape of their own, and do not resist changes of shape, as solids do. Their only elastic resistance is to compression, and this is so high as to lead one to say, for instance, that water is incompressible. This is not actually true, but a large force is needed to produce even a small shrinkage in most liquids. Bridgman has compressed water into three-fourths of its usual volume by employing a pressure of 20,000 atmospheres (when, curiously enough, it has to be heated to keep it from freezing). At the greatest oceanic depths (about 6 miles) the pressure is only about 1000 atmospheres, and the change of volume of water at this pressure would be small, not more than one per cent of the ordinary volume.

The coefficient of compression is defined as in gases. If the volume is V, the shrinkage v, and the pressure producing it p, the coefficient of compression is pV/v. Values for this coefficient are given in Table VI. Though liquids do not follow the gas laws, there is a similarity on one point, namely that they resist a very quick compression a little more than a slow one; but the difference

is not large, as in gases; on the contrary it is so small as to be usually negligible.

TABLE VI

Elasticity or Compressibility of Liquids

Change of volume per centimeter for a change of pressure of 1,000,000 dynes per square centimeter. (Multiply by 1.013 to change to compressibility per atmosphere.)

Water	0.000049
Mercury	0.0000037
Ether	0.000145

Elasticity of solids. It is an interesting question to consider what the construction of a solid must be in order to enable it to regain its shape after it has been stretched, or bent, or twisted a little. We make the natural assumption that solids are made of very small particles, atoms or molecules, very close to one another, held together by forces whose actions resemble those of springs, capable of yielding a little in any direction, but also capable of restoring the body to its natural shape and size after the distorting forces are removed. These molecular forces may prove to be electrical, but we need not consider their nature more fully at this stage. If we overstrain these springs, we may disarrange them and the body may be permanently changed in shape, or even broken to pieces. As we know, materials of different sorts vary enormously in their properties in these respects; some, like lead, are soft, yield easily and change their shape permanently (Fig. 9–3); while others like glass are hard and yield very little without breaking. When the arrangements of the particles in all materials and the nature of the forces between them are known, it will be easy to see why each sort of material behaves as it does.

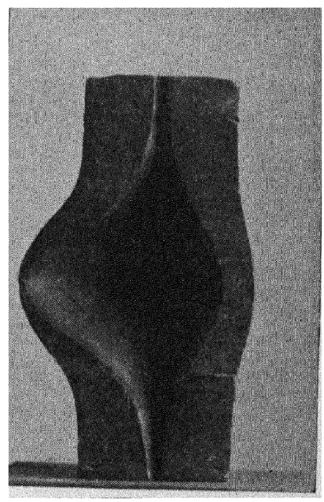


FIG. 9–3

A lead cylinder through which ran a $\frac{1}{4}$-inch hole. Professor Bridgman subjected it to a high internal pressure, and then it was cut in two.

Space Lattices. A good start toward discovering the reasons for the properties of common materials has been made through experiments with X-rays (Chap. 37). We now know how the atoms are arranged in all common crystals

such as rock salt, diamond, etc., and also in metals and alloys; and the influence of such operations as hardening, annealing, hammering, etc. on these arrangements has been investigated. Ordinary metals consist of a mixture of small crystals crowded together, oriented in every direction and of different sizes. Each crystal of any sort of material consists of an orderly arrangement of particles in rows and layers, piled on one another, which we now call a "space-lattice." Some of these are made up as though built of a multitude of cubes set close together. Others are like piles of balls, as in Fig. 9–4, in which the balls of the second layer settle down somewhat between the balls of the first and so on. This is called "close-packing" and is the arrangement in the case of gold, silver red-hot iron and other malleable metals. One can see that if the atoms are close-packed, one layer might be able to slide over another without separating the atoms much, and they might settle down into their new positions, having changed partners, but being as closely packed and therefore as "solid" as before. Hence such materials can be worked into new shapes without breaking. Other materials are made with such lattices that any stretching separates the atoms more and more, and these must be brittle substances.

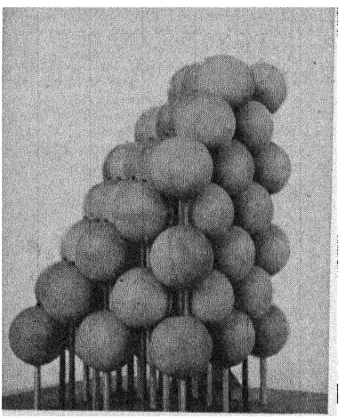


FIG. 9–4

A cut-away model of cubic close-packing ("face-centered cubic"). This arrangement enables planes in the crystal at several angles to slide over one another without separating.

The matter is complicated in the case of most metals by the fact that they contain crystals which are mixed together and may slide over one another when the material is strained. Thus the forces between crystals as well as those within them influence the properties of the material.

Hooke's law. Most bodies can yield a little and recover their original condition perfectly when the deforming force is removed. Within this limit the body is said to be elastic, and the effect produced is proportional to the cause producing it. This statement is known as Hooke's [1] law. In the case of a stretched wire, the extension is proportional to the stretching force, within the elastic limit. Similar statements can be made for bending, compression, etc.

[1] Robert Hooke, (1635–1703), English experimental physicist. He had many original ideas, but carried few of them to a successful completion. He partly anticipated Newton's law of gravitation and engaged in a controversy with him over priority. Though in many ways honored by his peers, he had similar experiences with others over optical and mechanical questions, and spent the latter part of his life as an embittered and penurious hermit.

It is often important to know how a material behaves when it is strained beyond its elastic limit, and what is its breaking strength. For these practical details, which vary from substance to substance the reader is referred to engineering handbooks, or other sets of tables.

Fundamental elastic deformations. Solids differ from liquids in having a definite shape. Not only may they be compressed, like liquids, but they may also be stretched, twisted, bent, and so forth. Of all possible elastic changes, only two are quite independent and different from each other. These are ***compression*** without change of shape, and ***shear,*** which is change of shape without change of volume. If we take a cube of jelly, we can easily distort it from the shape $BACD$ to the shape $BAEF$, by means of a force applied along the top surface, as represented in Fig. 9-5. This sort of change is an example of a shear. There is a ***shear coefficient*** (n) for solids, as well as the compression, or volume coefficient (k), as stated for liquids (p. 120). The shear coefficient is equal to the shearing "stress" divided by the "strain" which is

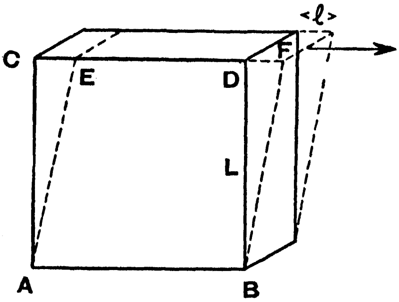


Fig. 9-5
A shear

$$n = \frac{\text{force produced per unit area of face}}{\text{shear produced per unit height}} = \frac{F/A}{l/L} = \frac{FL}{lA}.$$

This coefficient is also known as the "rigidity."

Other elastic coefficients. Many elastic changes involve both of the fundamental sorts, compression and shear; for instance, when a wire is stretched, its material is expanded a little (negative compression) and at the same time its layers flow over one another (shear). Bending a rod expands the convex side, compresses the concave side, and produces shearing of layer over layer, except along the middle of the rod.

Torsion (twisting) in a rod or wire is a shearing action. It can be shown that the torque L required to twist a wire of length l and radius r through a small angle θ is equal to

$$L = \pi r^4\theta\, n/2l.$$

The torque thus depends on one elastic coefficient only, the rigidity, or shear coefficient. This formula is useful in cases like the Cavendish gravitational experiment (p. 110), where a force is to be measured from the angle of twist of a wire.

Other elastic changes have coefficients of their own, but it can be shown that these are always expressible in terms of the two fundamental ones, the compression and shear coefficients. Only one more will be mentioned, namely the ***stretch coefficient,*** or "Young's modulus," which can be proved to be equal to $k + 4n/3$. It is defined as

$$Y = \frac{\text{stretching force per unit area of cross-section}}{\text{stretch per unit length}}$$

$$= \frac{F/a}{l/L} = \frac{FL}{al}$$

where l is the extension, L the whole length, F the stretching force and a the area. This coefficient is very useful in practical cases dealing with the stretching of wires, the compression of rods, and so forth.

TABLE VII

Elastic Coefficients of Solids

To convert Y from dynes/sq. cm. into lbs./sq. in. divide by 67,600.

	Young's coefficient Y in dynes/cm.^{2}	Rigidity or shear coefficient n in dynes/cm.^{2}	Volume coefficient k in dynes/cm.^{2}
Steel	20.9×10^{11}	8.12×10^{11}	16.4×10^{11}
Wrought Iron	19.20 "	7–8 "	14.6 "
Copper	12.5 "	4 "	14.3 "
Fused quartz	5.2 "	3 "	1.4 "
Duralumin	7.4 "		

All elastic coefficients are very large numbers, Y, for instance, being equal to the force which would (if such things were possible) stretch a one-centimeter cube of a material out to double its length, as the formula shows, if we take L, a and l as unity. They may be expressed in a variety of units. The elastic coefficients of solids are diminished by rise of temperature; thus, spring balances stretch a little farther for the same weight when hot than when cold.

Spring and torsion balances. In the ordinary spring balance a helical spring has a hook and a pointer attached to its lower end

which plays over a scale, usually uniform. The extension of the spring is roughly proportional to the stretching weight. In a more accurate form (Fig. 9–6) the spring is wound in a slightly conical shape, of smaller diameter above than below, so that the stiffness is greater above, where the weight of the spring itself has to be borne. The distortion in such a spring is more nearly uniform, and quite accurate results can be obtained with it. All spring balances suffer small changes with age and with temperature.

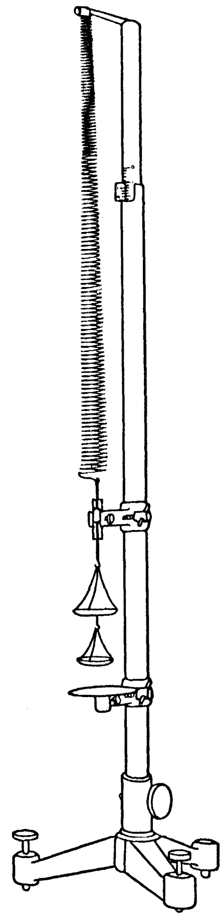


FIG. 9–6
A good spring balance

Torsion balances are of several forms, in which the twisting of an elastic strip or wire is made to oppose the forces to be measured or to indicate small changes in them.

The range of elastic forces. When a china plate is broken in two the pieces may be fitted very closely together, and yet they do not cling perceptibly to each other. Accurately flat glass surfaces do, however, begin to show forces of attraction for each other when pressed together in dust-free air; but even the best flat surfaces are not accurately flat in a molecular sense. They make contact at a few points only. These facts show that the forces of attraction that hold the particles of a solid together are felt through very small distances only, distances that prove to be of the order of a millionth of a centimeter or less.

The crystalline structure of metals. By etching a polished surface we can usually see that a metal is composed of a mixture of small crystals of different sizes crowded together and oriented in every direction. Figure 9–7 shows a microphotograph of such a surface. The properties of metals depend largely on the arrangement and size of these crystals, as well as the material which connects them. Annealing and hardening results in altering the size of the crystals. A long heating at a high temperature gives them time to grow together into larger aggregates. Finer crystals as a rule indicate hardness. Impurities in the metal produce their greatest effect

by altering the material acting as a sort of cement between the crystals; thus a small amount of carbon mixed with iron profoundly alters its physical properties, and turns it into steel.

It is now possible to "grow" large single crystals of a pure metal. These have properties which differ from those of the metal in its ordinary state, and their elasticity is different in different directions, as in crystals of other materials, such as quartz, for instance.

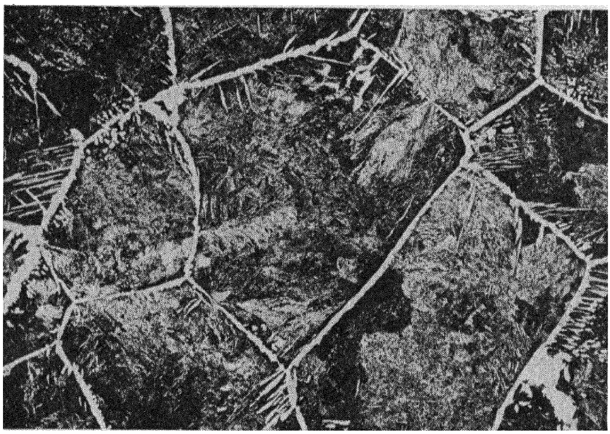


FIG. 9-7

An etched surface of cast steel, here magnified about 35 diameters. The large crystals (sorbito-pearlite) contain 0.5% carbon, the connecting material being simply iron (ferrite). Photograph by E. L. Reed, Harvard University,

Elastic lag. Ordinary substances are found to be far from simple in their elastic behavior when they are critically examined. One common defect is a tendency shared to a small extent by a wide variety of metals, which goes by the name of elastic lag. It can be shown in an exaggerated way by using an ordinary soft rubber band and hanging a weight on it. The rubber will be stretched out to a certain length at once, and this extension will slowly grow for a long time afterwards. Likewise, when the weight is removed, the rubber will not immediately shrink to its former length, but will gradually approach it. Glass sometimes shows this effect in thermometers.

PROBLEMS

1. A long glass tube closed at its upper end, is lowered into deep water. If the water at the bottom compresses the air in the tube to one-tenth of its previous volume, what is the depth of the water?

2. A vacuum pump has a cylinder of such size that each stroke of the piston creates an air chamber within the cylinder of 200 cc. volume, and the latter part of each stroke completely empties this chamber. If the pump is connected

to a bottle of 500 cc. volume, full of air at atmospheric pressure, what will the pressure in the bottle be after two complete strokes of the pump? After ten strokes?

3. If the value of Young's coefficient for steel is 2,200,000 kg./cm.^{2}, how much will a steel bar 1 cm. in diameter and 5 m. long stretch when a weight of 500 kg. is hung on it?

4. Calculate the stretch coefficient of a wire, if its length is 10 m., diameter 1 mm., the stretching force 10 kg., and the elongation 1 cm.

5. An air bubble is clinging to the bottom of a pail full of water, 25 cm. deep. Would any change in the volume of the bubble occur if the pail were allowed to fall freely?

6. Quoting from a recent newspaper article, we read that sunken ships going down in very deep water "come to a point beyond which their weight is less than the pressure of the water, and thus they are buoyed up." Criticize this statement, and show what factors are involved in determining whether a sinking object goes all the way to the bottom at great depths, or not.

7. Find the dimensions of the various elastic coefficients.

SURFACE TENSION

Cause of surface action. There are a number of familiar and very curious actions associated with liquid surfaces, which are explained as due to the same sort of molecular forces that produce elasticity. We must assume that in a liquid the molecules attract one another, when they come close enough, but that these forces are equal in amount all around each liquid molecule, so that they balance out and have no net resultant. Thus the particle is not held in any one position but is free to move in any way whatever, just as though there were no such forces. A particle at the surface, on the contrary, is not completely surrounded, and the net result of all attractions from neighboring particles is to pull it back into the liquid. This force rounds all corners, and if no other forces are acting, makes the liquid take a perfectly spherical shape, a form that is exhibited by drops of one liquid suspended in a liquid of its own density with which it does not mix.

This experiment is easily shown in the projection lantern, by filling a flat-sided cell with water, dropping a few salt crystals in the bottom to increase the density of the lowest layers, and a little alcohol at the top to produce lighter layers above; then by means of a glass tube, placing a drop of ortho-toluidine in the middle of this stratified liquid, which sinks till it reaches the layer

having its own density (Fig. 9–8). The use of previously boiled water will prevent the formation on the drop of little air-bubbles, which act disconcertingly like life preservers.

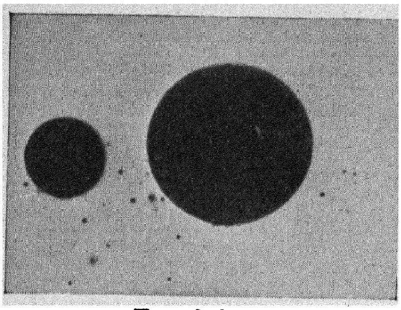


Fig. 9–8
Spherical drops resting in a liquid of their own density

Another interesting case is that of Fig. 9–9, which shows a fine stream of water in the act of breaking into drops, photographed by the light of a single, instantaneous electric spark. The stream is shown contracting into long narrow necks connecting larger amounts of water which presently become drops, oscillating wildly as they fall. The necks break into very small drops between the larger ones, which take a spherical form almost at once.[1]

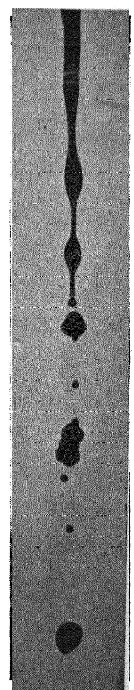


Fig. 9–9
A falling stream of water

Surface tension. Such behavior is commonly "explained" in terms of an imaginary but very convenient force of ***surface tension.*** The surface does tend to shrink as if under tension, but in reality this occurs on account of the attraction exerted on the surface particles by those below. A drop of water appears to be enclosed by a stretched elastic sheet, but, if this were really so, the tension would increase with the amount of stretching, according to Hooke's law. When, however, a liquid surface is stretched (as, for instance, the soap film shown in Fig. 9–10), the increase is made possible by the addition of new particles moving up from below, and since these are acted on by precisely the same forces as the original surface particles, ***the surface tension remains practically constant,*** independent of the extent of the surface.

It is easy to measure the amount of this surface tension approximately, though accurate measurements are difficult because of the relatively large influence of impurities, surface dirt, etc., the last traces of which

[1] We owe this interesting figure and the next to Messrs. D. B. Woodbridge and A. E. Parker, formerly students in Amherst College, carrying on this research under the direction of Professor S. R. Williams.

TABLE VIII

Surface Tensions of Various Liquids

(In dynes per centimeter)

Water at 15° C.	73
at 100° C.	58
Mercury at 17° C.	547
Liquid oxygen at − 183° C.	13
Acetone at 17° C.	23
Olive oil at 20° C.	32

are extremely hard to remove. In a good instrument for this purpose (Fig. 9-11) a thin platinum wire bent in the form of a circular ring is supported on a delicate torsion balance. Platinum is chosen as a material for the ring because it can be heated red-hot in an open flame, and so deprived of all traces of grease, always the worst enemy to overcome in such work. This ring, held horizontal throughout the experiment, when dipped into a clean water surface and pulled upward by twisting the torsion wire, brings with it a film of water clinging to both the inner and outer sides of the wire, as the figure shows, even when the wire has reached a height of several millimeters above the level of the surface. This film from the water up to the ring is then vertical, and has a total length of twice the circumference of the ring; twice, because the film is doubled. The force necessary to hold the ring in this position against the tension of the film can be measured from the angle through which the torsion wire is twisted, and is usually so small that the dyne becomes a convenient unit in which to express it. We define the surface tension as ***the force across a*** (linear) ***centimeter of the surface film.*** Its value is readily obtained from these measurements. (See Table VIII.)

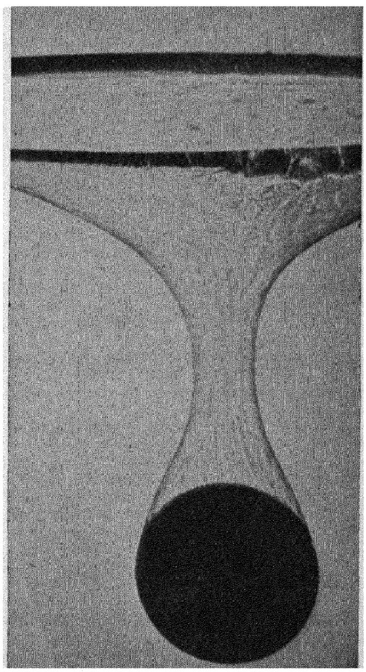


FIG. 9-10

A spark photograph of a ping-pong ball falling through a soap film

The spreading of oil over a surface. Oil floating on water furnishes some interesting phenomena. With many sorts of oil the first drop that touches the water spreads with great speed over the whole surface, covering it with an almost invisible film. If more oil is added it may stand up in the form of drops on the

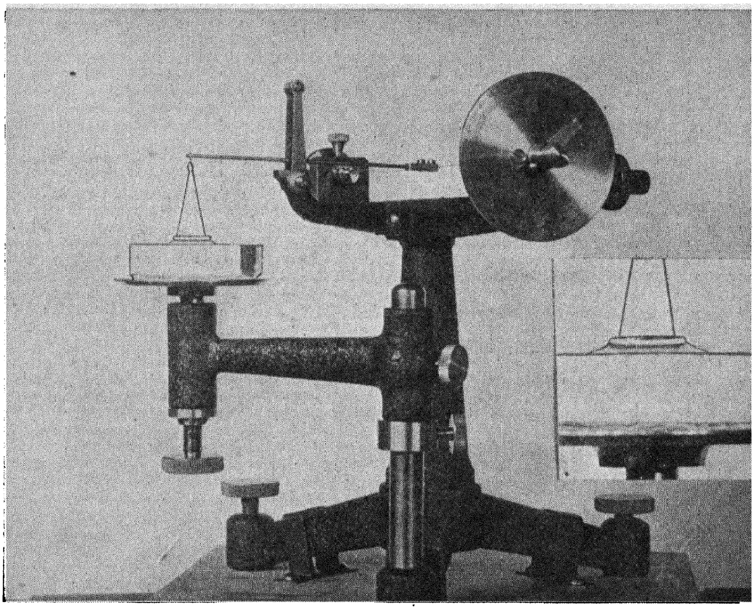


FIG. 9–11

A torsion balance for the measurement of surface tension. By turning the large dial, a horizontal wire is twisted, and this lifts the little ring out of the liquid. The insert shows the ring supporting a film of water.

first film, or it may spread uniformly. Whether a drop spreads or not depends on the surface tensions involved. A particle at the corner of a drop, like P in Fig. 9–12, is acted on by one force in the direction PS, and by other forces in the directions PQ and PR. The amounts of the forces PQ and PR depend on the nature of the oil. If the drop comes to rest in the form shown, the force PS must be equal and opposite to the resultant of PQ and PR. If PS were greater than the arithmetical sum of the other two, the drop would spread out over the whole surface, and this is what happens to the first drop on a fresh surface of water.

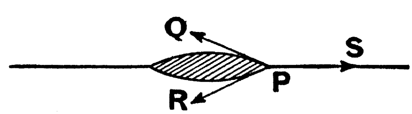


FIG. 9–12

If the water surface is of great extent, or the quantity of oil exceedingly small, the oil spreads out over the surface in a layer which is just one molecule thick. The existence of such a film can be shown by dusting the surface with lycopodium powder before adding the oil. The powder retires, leaving the oiled surface clear; Fig. 9–13 was prepared in this way.

Changes in surface tension. Various causes alter the surface tension of liquids. The effect of temperature has already been mentioned, and can be shown by a simple experiment. If we float two slivers of wood near together on water, and plunge a small hot body into the water between them, they will instantly spring apart, the unweakened film on the outside drawing them away. A similar experiment done by putting a glass rod moistened with alcohol between the bits of wood, shows the effect of dissolved substances, which also usually weaken surface tension. Dusting the surface with lycopodium or talcum powder shows beautifully just which part of the surface is affected in these experiments. It should be noted also that the surface tension of a liquid varies slightly with the gas or vapor with which it is in contact.

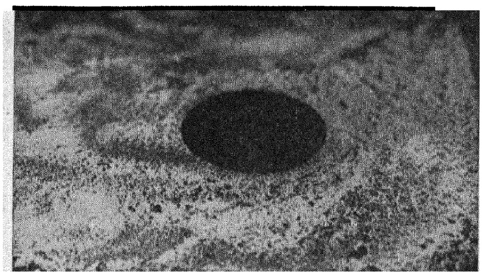


Fig. 9–13

Rayleigh's determination of the size of an oil molecule. An important conclusion can be drawn from a frivolous-looking experiment of this last sort. A few bits of camphor are dropped on a clean water surface. They spin and travel about in a most irregular manner for a very long time. The explanation of their behavior is that the camphor vapor dissolves slightly in the water, weakening the surface tension most at those places where this occurs most readily. As the pieces are irregular, the action is never uniform, and the camphor bits are pulled about, always in the direction of the stronger (less affected) film. A remarkable change occurs at the moment when a drop of olive oil is touched to the surface. It spreads almost instantly over the whole surface, weakening the surface tension as it goes, so that the camphor bits jump away from it when it touches the water and then immediately come to rest.

The oil film evidently prevents the camphor vapor from coming into contact with the water. A study of organic chemistry leads us to believe that these oil molecules are elongated in shape, with one end of a sort to be attracted to the water and held by it, while the other end is inactive. They therefore form a uniform film all over the water, each molecule standing on its head, so to speak, with its inactive tail in the air, and if there is not too much oil this film may be only one molecule thick. Less oil than this will leave open patches of water in which the camphor particles will still move; more oil than this will pile up on top of the uniform layer. From these considerations the late Lord Rayleigh[1] devised a simple experiment which yielded a good estimate of the length of one of these oil molecules. He used a large pan of known surface area, filled with fresh water, with bits of camphor twirling about on it. On this surface he dropped a small (and carefully weighed) drop of oil, and observed whether or not it stopped the camphor motions. By several trials, the smallest amount that would just stop the motions was found, and since the mass and density of the oil was known, the volume of this amount could be obtained. From this its thickness when spread over the area of the large pan could be estimated. The result was that the oil molecule was found to be 2×10^{-7} cm. long. It should be noted that this molecule is many times too small ever to be seen.

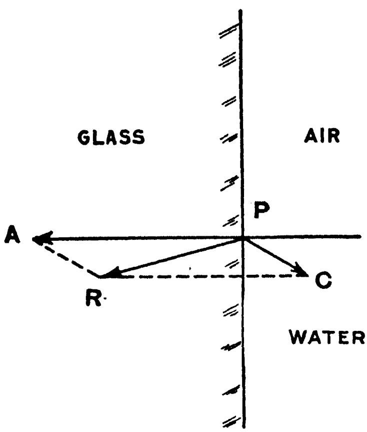


FIG. 9–14

Adhesion and cohesion. When two different materials (liquid and liquid, or liquid and solid) are in contact not only do the

[1] John William Strutt, third Lord Rayleigh (1842–1919), a brilliant mathematician and experimenter, whose original papers, collected in five large volumes, cover almost the whole range of physics. He discovered the inert gas *argon* in the air, received the Nobel prize in 1904 and occupied posts as Professor (later Chancellor) at the University of Cambridge, and later at the Royal Institution. He was a master in the art of obtaining accurate results of fundamental importance with modest experimental facilities. His son, the present Lord Rayleigh, has worked in some of the same fields with almost equal distinction.

molecular forces already spoken of tend to keep like molecules together, but there are also forces of attraction between unlike particles. The former action is called ***cohesion,*** the latter ***adhesion.*** As an example let us consider glass and water in contact, on the inside of a drinking glass. If the water came up quite level to the very corner, as at P, Fig. 9–14, next to the glass, there would be a force of cohesion PC drawing the corner particle down into the water, and a force of adhesion PA drawing it toward the glass. With clean glass, PA is larger than PC, as drawn in the figure. The resultant of these forces is PR. The surface of a liquid at rest always sets itself at right angles to the resultant force. It cannot, therefore, remain as in Fig. 9–14, but will take the form shown in Fig. 9–15, running up the side of the glass. As it does so, the particles next to the glass are acted on by a cohesive force which becomes more nearly straight down, so that the resultant PR becomes nearly perpendicular to the glass. This allows the water to cling at a considerable height, especially if it is put into a narrow ("capillary") tube. The rise of liquids in such experiments is often said to be due to ***"capillarity."***

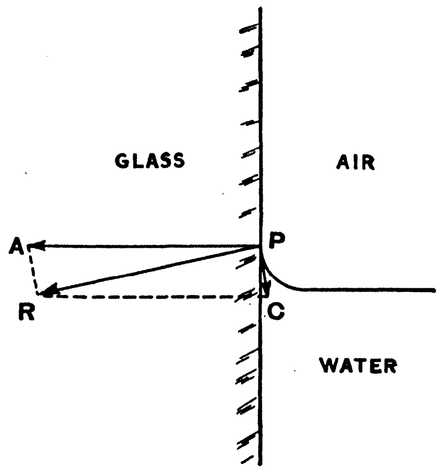


FIG. 9–15

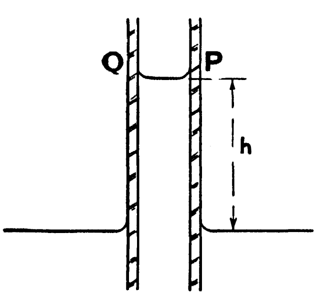


FIG. 9–16

Rise of liquids in tubes. The height to which a liquid will rise in a tube can easily be calculated. Figure 9–16 represents a section of a glass tube dipping into a liquid. Around the tube in a ring at PQ the liquid is clinging to the glass with its surface there practically vertical. The surface tension is T dynes across every centimeter of film, or $2\pi rT$ dynes in all, if r is the radius of the "bore" of the tube. This force supports a cylinder of liquid h cm. high, which

weighs $h \times \pi r^2 dg$, d being the density, and g being introduced in order to put the weight into dynes. Hence

$$2\pi r T = h \times \pi r^2 dg,$$

or

$$h = \frac{2T}{rdg}.$$

This arrangement may be used to measure the surface tension of the liquid. If the tube is round, and its diameter is carefully measured under a microscope, and ***if everything is clean,*** good values may be obtained. It is best to use a narrow tube in order to get an easily measurable height, and such tubes are often not round, and are hard to measure in cross section. Surface tension experiments are so delicate that a fresh clean surface will become contaminated merely by being exposed to the air of an ordinary room for a few minutes.

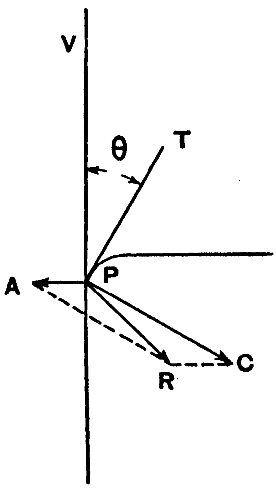


FIG. 9–17

Depression in tubes. In the case of water in contact with paraffin, or mercury with glass, the cohesive force is so much greater than the adhesive that the liquid does not run up the solid and "wet" it, but stands away from it with the edge of the surface running down. The forces (Fig. 9–17) for this case show a new condition. The resultant PR does not become horizontal, but stands at a definite angle. This means, if P is, as before, the "corner particle," that the liquid meets the surface at a definite angle θ in the figure, called the angle of contact. In calculating the depression of the surface in a glass tube partly filled with mercury, we have as before the weight of a cylinder, $\pi r^2 hdg$ dynes, acting in this case upward, and held down by a surface tension force $2\pi r T \cos\theta$, since the tension is along TP and its vertical component, along VP, is the only part of the force which is useful in depressing the mercury. Hence

$$T = \frac{hdgr}{2\cos\theta} \quad \text{or} \quad h = \frac{2T\cos\theta}{rdg},$$

and this cannot be solved without first finding the observed value of the angle of contact. For mercury and ***clean*** glass θ is about

40°; unfortunately glass is usually not clean, and the value then becomes a little uncertain.

The depression of the surface is easily shown (Fig. 9–18) by putting some mercury into a glass U-tube with arms of very unequal internal diameters. If one is very large, say 2 cm., the depression on that side is practically negligible; if the arms were of equal diameter, it would be the same on both sides, that is it would not occur at all because the depressing forces would be balanced. This depression is important in the use of barometers, accurate thermometers, etc., since it must be allowed for as a correction.

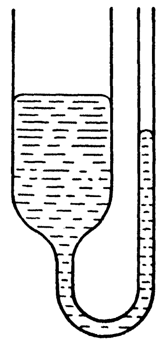


Fig. 9–18

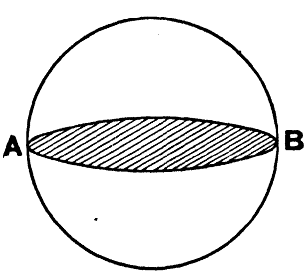


Fig. 9–19

Pressure in drops. A drop of liquid in a spherical form, or a soap bubble, has a pressure inside, due to surface tension. Consider a drop cut in two by an imaginary plane. The surface tension holding the halves together is that acting in the film (Fig. 9–19) across the circumference of the circle AB, which is $2\pi r T$, if r is the radius. The internal pressure p multiplied by the area of the circle AB must balance this force.

Hence $$\pi r^2 p = 2\pi r T \quad \text{or} \quad p = \frac{2T}{r}.$$

The pressure inside a soap bubble will be twice as much for the reason that there are two surfaces, the inner and the outer, each under the tension T.

Tension inside a liquid. When water rises in a capillary tube, the column which is above the normal level is held up by surface tension, and by adhesion to the wall of the tube. The liquid inside must be in a ***state of tension.*** The forces of attraction between the particles must allow the column of liquid to hang, as it were, from the surface and thus from the wall of the tube. One cannot ordinarily grasp the top of a quantity of liquid and pull it up because the liquid breaks up into drops with air between, and falls back. Here it is prevented from doing so by the walls of the tube, which keep the air out.

A similar exhibition of tension in a liquid may arise in a clean tube which is filled with mercury, as in the construction of the mercury barometer (p. 16). If very great precautions are taken

to ensure cleanliness and avoid air bubbles, the mercury may not fall as it usually does when the barometer tube is set vertical, and in this way a column of mercury of double the usual height has been supported. In this case the mercury was supporting a tension of at least 15 lbs./in.^{2}, and the adhesive force at the top of the tube was doing the same.

A small drop of water between two large pieces of plate glass will hold them together with a very great force. If we try to separate them the water takes the form of a circular disc, the outer edge of which is grooved like the wheel of a pulley. The surface tension at this surface tends to draw the two plates together, and also to straighten out the edge, and thus to stretch the water out into a still larger drop. This puts the water into a state of tension, and the value of this force can be calculated and also directly measured. All computations of the tensile strength of water indicate a surprisingly large value, not far from that of steel. This is really the same as the cohesion of water, since the cohesion is what prevents it from stretching. A large force of cohesion also implies a large ***internal pressure*** in the liquid, of whose existence we are usually unaware. Those liquids in which this internal pressure has a large value should be more difficult to compress. They should also have a high surface tension. Hence one might expect the compressibility of a liquid to vary more or less inversely as the surface tension. This relation is not exactly followed; but in many cases is approximately true.

Pseudo-biological actions. A number of curious experiments can be performed (best in the projection lantern) in which solids and liquids, acting under the influence of the forces considered in this chapter, move about almost as though they were alive. One of these will be cited. On a water surface of 3 or 4 in. diameter a few scraps of camphor are placed, which proceed to move actively about. If a few drops of ortho-toluidine (which we shall call "the red liquid") are added, the camphor motions usually cease, and the particles become "dead." The drops themselves flatten out somewhat and appear to dissolve irregularly and actively in little filaments at the edges, until all of a sudden an unstable condition is reached, when an explosive action occurs, and the drop writhes about with violent contortions. A speck of camphor dropped into the middle of one of the drops of red liquid will blow it to pieces, and thereafter act like a demon, rushing madly over the surface, pursuing red liquid wherever found, and apparently trying to break it up. It seems very much alive compared to the "dead" particles already there, and it retains its activity for a minute or longer. If now a drop of a second liquid, di-methyl-aniline, is placed on the water, it will sit quietly on the surface in a sharply

bounded drop. A scrap of camphor and the drop of the second liquid are attracted to each other; the liquid will "attack" the camphor, "swallow" it and dissolve it. It is easier in the case of several of these experiments to describe them than to explain them. As a class they suggest the possibility that some, at least, of the motions of low forms of living creatures may be due to surface tension.

If a drop of mercury is placed in a dilute nitric acid solution in a flat dish, and a crystal of potassium bichromate is placed near it, the mercury moves about in a manner resembling the simple animals called by biologists "amœbæ."

Books recommended:

C. V. Boys "Soap bubbles."

A. S. C. Lawrence, "Soap Films," 1929 (G. Bell and Sons)

PROBLEMS

1. What is the force acting on a match 5 cm. long, floating on water, if a drop of alcohol weakens the surface tension on one side of the match from its usual value of 73 dynes to a value of 53? Which way does the match move?

2. A moth floating on the surface of water is wet by it to such an extent that to get out of the water it must have to break a total length of 6 cm. of (single) water film. If the moth weighs 1 gram, how does the force of surface tension holding it in the water compare with its weight?

3. Find the height to which a liquid of density 0.8 and surface tension 30 dynes would rise in a tube of 0.6 mm. internal diameter? (Angle of contact = 0.)

4. Two soap bubbles of unequal size are blown one on each end of a tube, with a free connection which can be opened between them. What will happen when this is done?

5. Find the pressure (in grams/cm.^{2}) produced in a drop of water of 0.00005 cm. diameter (which is about as small as can be seen under a strong microscope).

6. Find the internal diameter of a glass capillary tube in which water rises to a height of 10 cm.

7. Hot paraffin spilled on a woolen coat can be taken out by covering the spot with blotting paper and applying a hot iron. Why should the paraffin run into the blotting paper instead of spreading over a larger area in the coat?

8. When the soil is fairly compact water can come up to the surface and be lost by evaporation, but if the surface layer is broken up and pulverized most of the water can be retained. Explain.

9. There are insects that can run along on the surface of water. Does Archimedes' principle apply to them, as floating bodies?

10. Snails sometimes come to the surface of water and travel along it on the under side of the surface film. One sometimes sees the statement that they are clinging to the film, and that this is partly supporting them. Can this be true, considering the cause of surface tension?

CHAPTER 10

KINETIC THEORY OF GASES

First assumptions, 138; motion of gas particles, 139; the Brownian movement, 139; the motions of evaporating particles, 140; the diffusion or condensation pump, 141; the average velocity of gas particles, 141; the nature and amount of gas pressure, 142; the actual velocity of the particles, 143; kinetic theory and the laws of gases, 143; the lowest temperature, 144; kinetic theory of liquids and solids, 144; Avogadro's rule, 144; relations among molecular quantities, 145; gas viscosity and the mean free path, 145; molecular radius, 145; molecular dimensions, 146.

Any good physical theory starts from a few simple and likely hypotheses and deduces from these a number of conclusions capable of experimental test, and of explaining a large and varied series of observed facts. The kinetic theory of gases is one of the most interesting of these, as its hypotheses and reasoning are simple and its results far-reaching, often surprising, but always very near to the truth.

First assumptions. We begin by considering how very easily gases can be compressed. Common experience shows that a quantity of gas can be forced without too much trouble into a space many times smaller than that which it usually occupies. This leads us to the natural assumption that ordinary ***gases consist largely of empty space.*** Then we must suppose that there is something that the substance of the gas is made of, and we assume this to take the form of ***small particles.***[1] We must also admit that these are in motion, for we observe that odors spread throughout a room or a house by the process which we call ***diffusion.*** Of course, convection (p. 177) is the quickest way of making gases mix with each other, but diffusion alone will do it in time, if convection is prevented. We know, too, that if we break an incandescent electric lamp bulb which has a vacuum inside, the air rushes in with almost explosive energy; also, in a vacuum

[1] The smallest particles are called atoms by the chemist, or if small groups of atoms are firmly bound together, they are then called molecules. In this discussion we shall avoid such distinctions by referring to them as particles throughout.

pump with moving piston the air follows the receding piston without delay; or, when an automobile tire blows out, it takes a very short time for the air in it to escape. These illustrations, and many more, are in harmony with the idea that ***the gas particles are in rapid motion.***

Motion of gas particles. Such particles must, of course, collide with one another and with the walls of the containing vessel very frequently. If they were tennis balls rushing madly about inside a large gymnasium, let us say, instead of gas particles in a bottle, we should expect soon to find them all resting on the floor. We observe that air does not settle on the bottom of a bottle in this way, leaving a vacuum above; so we must suppose that the particles do not stop moving and that the collisions do not involve any loss of kinetic energy. This may seem more reasonable if we imagine that the particles do not actually come into contact at all at a collision, but begin to repel each other from a little distance and rebound without touching; but the range of this repulsive force must be extremely small. Also, we shall suppose for the present that no other force acts between the particles at any distance. These assumptions make it easy to see the difference between gas particles and tennis balls. If the motions and the collisions of the latter were frictionless, they also would rebound forever inside the gymnasium.

The Brownian movement. The agitation of the particles of a gas can almost be seen in a microscope; we say ***almost*** because, although the particles themselves are invisible, their effects are not. An astonishing phenomenon was discovered in 1827, long before the days of the kinetic theory, by Brown [1] and is now known as the Brownian movement. It can be observed in gases, though the experiment is more easily done with liquids. If a small quantity of an insoluble substance (powdered gamboge, rutile or titanium oxide, or finely ground mica are recommended) is mixed with water, it forms a fluid containing a mixture of solid particles of many sizes. When a drop of this is viewed under a good microscope, magnifying five hundred times or more, the smaller particles are seen to be in a state of continual, irregular agitation. The smallest particles have the liveliest motions. The very large ones remain at rest. The explanation of this curious effect is simple. The visible particles are all extremely large compared with the

[1] Robert Brown (1773–1858), an English botanist.

particles of the liquid itself and are continually being struck by them. If the large ones have so great a surface that at every instant of time they are receiving myriads of blows on all sides, these blows will perfectly balance, and no motion will be caused by them. But if the solid particles are, say, only a thousand times as heavy as the moving particles of the liquid, the blows will not at all times be symmetrically distributed, since this distribution is an effect of chance, and at one instant there will be more blows on some one side than on the opposite. Hence the solid particle will tend to spin around or to move over a little in the direction demanded by the greater number of blows. Evidently this lack of symmetry in the blows is the more likely to occur the smaller the solid particle is. The smallest ones that we can see in a good microscope move almost as though they were alive.

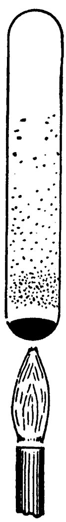


Fig. 10–1

To show the effect with a gas one needs a small shallow chamber with glass walls to put under the microscope, into which smoke can be drawn. The smoke particles should be seen by reflected light against a dark ground. As they are free to wander out of the field of view, the observations are a little more difficult, but the nature of the phenomenon is the same in both liquids and gases.

The motions of evaporating particles. A somewhat similar experiment which is easily projected on a lecture-room screen was recently devised by Knipp.[1] It exhibits effects due to the evaporation of particles (in this case atoms) from a surface of liquid mercury in a vacuum, and gives a most vivid picture of their activity.

A closed glass tube (Fig. 10–1) has a little pool of liquid mercury in the bottom, and above this a high vacuum. Floating on top of the mercury surface are a few small grains of broken glass. When the mercury is gently heated (at the bottom only) so as to encourage evaporation, the escaping particles going off in the form of mercury vapor rush upwards with such vigor that they drive the glass particles before them, and by a continual bombardment from below prevent them from returning, keeping them in a state of the wildest agitation. The evaporating mercury particles recondense into droplets on the upper, unheated part of the tube. Such a vapor as exists in this tube, escaping from the mercury surface, is unusual in that

[1] C. T. Knipp, Professor of Physics at the University of Illinois.

the motion of its particles is practically all in one direction, instead of being distributed at random.

The diffusion or condensation pump. The action just described has been turned to good account in the construction of a powerful vacuum pump known as the ***diffusion pump,*** which is now commonly used in the production of the highest vacua. In the form shown in Fig. 10–2, mercury is heated electrically, or by a small flame, and the evaporating atoms rush by the opening A on their way toward the cooler parts of the tube, where they are cooled by running water, condense into little drops, and fall into the liquid below. As the mercury particles rush by the opening, gas particles coming from the vessel to be exhausted get entangled in the stream, and are swept along, with no chance of returning. The gas thus removed from the vessel collects in the lower part of the pump B at a relatively high pressure (though still much below ordinary pressures), and from there an ordinary (usually oil) vacuum pump removes it. The diffusion pump thus operates with no moving parts larger than gas atoms or molecules. It is capable of producing a pressure as low as 0.000005 mm. of mercury, and is extensively used in the manufacture of incandescent lamps, X-ray and radio tubes, etc. Quite recently certain organic liquids have been found to be as successful in this pump as mercury, or perhaps even better.

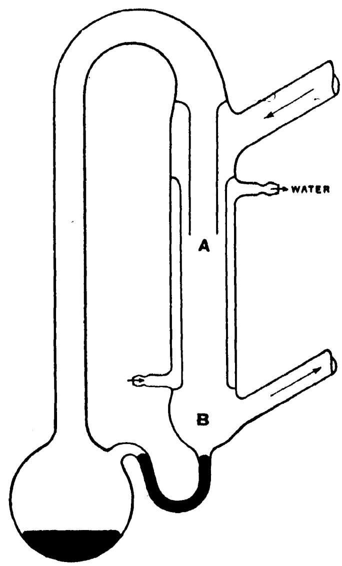


Fig. 10–2
A diffusion or condensation pump

The average velocity of gas particles. All the particles of a gas could not be moving with the same speed. If such an arrangement existed for one instant, the first collision between particles would upset it. As a matter of fact, some particles are likely at any given instant to be moving quite slowly, and others very fast, depending on the nature of the collisions they have recently experienced, and thus the distribution of speeds among the particles must be according to the laws of chance. There is, however, an average speed; in fact, it is possible to consider more than one

sort of average speed. The one we shall find most useful is the one which corresponds to the average kinetic energy of the particles; since this depends on the square of the speed, rather than the speed itself, our sort of average speed is really the square root of the average value of the square of the speed, or $\sqrt{\text{average (speed)}^2}$. This is sometimes referred to as the R.M.S. value (root mean square), an expression that will be of use later (Chap. 26).

In general the direction of the motion of the particles is perfectly random. If their velocities could be observed individually, it would be possible to resolve them into components along three reference axes mutually perpendicular to one another, and then we should find that on the average these components were equal; that is, there are equal amounts of velocity up or down, or right or left. A similar remark applies to the momentum.

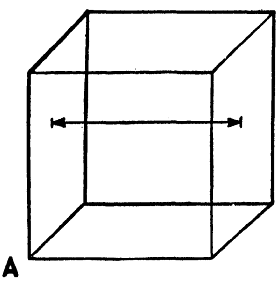


FIG. 10–3

The nature and amount of gas pressure. A large number of particles rebounding continually against a surface furnish a practically steady push on that surface, a fact which yields us the first deduction from the kinetic theory, namely an ***explanation of gas pressures.*** A similar problem is furnished by the calculation of the pressure exerted on a wall when a powerful stream of water is sent against it from a fire hose. We shall now calculate the amount of the pressure of a gas. Let p be the force per square centimeter, V the "average" velocity of the particles, as above defined, m the mass of each particle and N the number of particles per cubic centimeter. Let us for simplicity consider a cubical vessel, one centimeter side, (Fig. 10–3), filled with the gas and, instead of imagining the particles to be moving at random in all directions, let us suppose that $N/3$ particles are moving parallel to each of the three edges of the cube which meet at A, and all moving with velocity V.[1] Each particle moves forward 1 centimeter, collides with a wall, moves back 1 centimeter, forward again 1 centimeter, colliding again with the same wall, and so on. In other words, it moves 2 centimeters between collisions on the same wall. Doing this at a velocity V one particle collides $V/2$ times

[1] The real state of affairs is much more complex, but this simplification leads to essentially the same results.

a second against a particular wall, so that there are in all $N/3 \times V/2$ collisions with that wall per second. At each collision the forward momentum of a particle mV, is changed to backward momentum, or $-mV$. The momentum change at each collision is thus $2mV$. Hence the total change of momentum per second at one face is $N/3 \times V/2 \times 2mV = NmV^2/3$. But Newton's second law of motion defines the rate of change of momentum as the force; if m is in grams-mass, and V in centimeters a second, the force will be expressed in dynes (p. 66). This is the force against an area of 1 square centimeter, which is therefore the pressure on that face, or on any other, or on the same area in a larger vessel full of the same gas. Hence $p = NmV^2/3$ dynes per square centimeter. But Nm is the mass of material in a cubic centimeter, which is the density d, of the gas. Hence $p = dV^2/3$.

The actual velocity of the particles. In the formula just derived the pressure and the density are well-known quantities. Hence the velocity which we must imagine the particles to possess can be found. For hydrogen at standard temperature and pressure $d = 0.0000899$ grams/cm.^{3} and $p = 76 \times 13.6 \times 980 = 1{,}013{,}000$ dynes/cm.^{2} Hence $V = 1838$ meters a second. This is a little over a mile a second, a startlingly high velocity. According to the formula, the particles of denser gases must move less fast, the speed varying as the square root of the density, if the pressure is constant. Hence the particles of nitrogen (density 0.00125) and of oxygen (density 0.00143) which are the main constituents of air, move at a speed of about one-third of a mile a second.

Kinetic theory and the laws of gases. Boyle's law states that if v is the volume of a certain mass of gas, and p its pressure, the quantity pv is a constant, provided we keep the temperature from changing. Our pressure formula just deduced bears a striking resemblance to Boyle's law if we write it in the form $pv = MV^2/3$, putting M/v for d, M being the mass of the gas. In the latter form we see that the right-hand expression is proportional to the kinetic energy [1] of all the gas particles together. Experiments to

[1] The kinetic energy here referred to is that due to the motion of each particle as a whole, that is, the *kinetic energy of translation*. There may be other sorts of kinetic energy present, due to spin or to inner vibrations, but without going farther into this subject, we assume that the total kinetic energy will be proportionately divided among all the possible sorts, in which case the kinetic energy of translation will be proportional to the total kinetic energy, and we need not be very particular about mentioning one sort only.

be described in the next chapter (p. 160) show that the product pv is proportional to what we shall call the absolute temperature. Hence the kinetic theory will agree with the results of these experiments if we imagine that the absolute temperature is somehow connected with the kinetic energy of the gas. It is most convenient to assume that ***the absolute temperature is directly proportional to the average kinetic energy of the individual gas particles.*** This idea will be found very useful.

The lowest temperature. The temperature at which the gas particles have no kinetic energy at all would, according to the discussion above, be the lowest conceivable temperature. This we may call the ***absolute zero*** of the temperature scale. The existence of such a point is indicated by other considerations (p. 158).

Kinetic theory of liquids and solids. We have seen that in the Brownian movement we have experimental evidence of the kinetic agitation of the particles of a liquid. It is interesting to observe that the particles of any liquid or solid in contact with a gas ought to share the motion possessed by the gas particles. If, for example, the particles of a gas in a glass bottle were striking the particles of the glass and the latter were not in motion already, they would begin to move under the action of this bombardment, absorbing energy from the gas and thus cooling it. In any actual case, a transfer of energy does take place between the particles of a solid (or a liquid) and a gas in contact with it, until a state of equilibrium is reached, after which they exchange equally with each other. In terms of the language of heat, we describe this by saying that they come to the same temperature. It is thus a consequence of the kinetic theory of gases that we must also have a kinetic theory of liquids and of solids.

Avogadro's rule. Another triumph of the theory follows at once if we consider two gases of different kinds at the same pressure and temperature. Let the first gas have N_1 particles per cubic centimeter, each of mass m_1 and velocity V_1; while the second has similar quantities N_2, m_2 and V_2. Since the pressures are the same,

$$p = \frac{N_1 m_1 V_1^2}{3} = \frac{N_2 m_2 V_2^2}{3}.$$

Since the temperatures are the same, the average kinetic energies of the particles are equal, or

$$\frac{m_1 V_1^2}{2} = \frac{m_2 V_2^2}{2}.$$

Dividing one of these equations by the other, we derive $N_1 = N_2$; or, in other words, all gases at the same pressure and temperature contain the same number of particles per cubic centimeter. This is familiar to chemists under the name of ***Avogadro's*** [1] ***rule.***

Relations among molecular quantities. It is of interest to follow the kinetic theory through to the point of obtaining the actual number of particles per cubic centimeter in a gas, the mass and size of each, the average distance they travel between collisions, (called the mean free path), and the number of collisions they make per second. Evidently these quantities are connected with one another. For example, $Nm = d$, and if the number of particles per cubic centimeter can be found, the mass of each will be disclosed. Also, if their mean free path is l, their velocity V and the number of collisions they make per second is q, q must be the number of times they travel a distance l in a second, or $q = V/l$. The number N and the volume of each particle (which we may call v') are connected, for their product Nv' must be the total volume of the actual material in the gas, which we can roughly estimate by liquefying the gas, and observing the volume of the liquid. This is equivalent to assuming that the particles are practically "in contact" with no waste space between them, when in the liquid form. The high resistance to compression offered by liquids makes this seem probable, and other evidence also points to the view that the empty space in a liquid may be only two or three times that occupied by the particles themselves, thus justifying the above assumption.

Gas viscosity and the mean free path. A definite value for the mean free path comes out of the study of viscosity in gases. We shall not follow this through in detail, but it is interesting to see how the kinetic theory accounts for friction in a gas. Friction would be expected to occur when one layer of a gas moves over another. In addition to the random agitation of its particles which all share, the moving layer possesses an average forward motion. The random motion of the particles near the plane separating the two layers will carry a great many particles from the moving layer into the other, and also in the opposite direction. This will bring part of the forward drift of the moving layer into the stationary one, and conversely; so that the action is just like a frictional effect, dragging back the moving layer and dragging forward the stationary one. A detailed study of this effect, together with measurements of the viscosity of gases leads to a value for the mean free path of 1.83×10^{-5} cm. for hydrogen under standard conditions of pressure and temperature. For other common gases it is two or three times smaller.

It is a direct and unexpected consequence of the kinetic theory that the viscosity of gases ought not to depend on the pressure. This would be absurd, of course, if the pressure became nearly zero, because there could be no friction if there were no gas; but, within wide limits, and down to a pressure of about one-hundredth of that of the atmosphere this conclusion is experimentally verified.

Molecular radius. A gas particle rushing in among others (supposed for convenience to be at rest) will sooner or later hit one of them. The distance

[1] Count Amedeo Avogadro (1776–1856), Professor at Turin University.

TABLE IX

Molecular Quantities at 0° C. and Standard Pressure

	Hydrogen	Oxygen	Carbon dioxide
Number of particles per cm.^{3}	2.705×10^{19}	same	same
Diameter of each (molecule)	2.4×10^{-8} cm.	3.2×10^{-8} cm.	4.2×10^{-8} cm.
Mass of each molecule	3.33×10^{-24} gm.	53×10^{-24} gm.	93×10^{-24} gm.
Mean free path	1.83×10^{-5} cm.	1.0×10^{-5} cm.	0.62×10^{-5} cm.
Number of collisions per sec.	1.00×10^{10}	4.6×10^{9}	6.2×10^{9}
Average velocity	18.4×10^{4} cm./sec.	4.6×10^{4} cm./sec.	3.92×10^{4} cm./sec.
1 cubic centimeter weighs	8.99×10^{-5} gm.	1.43×10^{-3} gm.	1.98×10^{-3} gm.
1 gram occupies	1.112×10^{4} cm.^{3}	699 cm.^{3}	505 cm.^{3}
Number of molecules in m grams (m = molecular weight)[a]	6.06×10^{23}	same	same

[a] This number is the same for all substances whether gaseous, liquid or solid.

which it has to go (on the average) before doing so is the "mean free path" and this depends on how many particles there are, and how large they are; that is, on how much area they appear to cover as seen from the direction of the entering particle. Simple reasoning on this basis leads to the value for the mean free path, $l = \frac{1}{\pi r^2 N}$, where N is as usual the number of particles per centimeter and r is the radius of each. This result will not be materially altered if the particles are all in motion. We have already arrived at an estimate of the total volume of all the material in 1 cm.^{3} of the gas. The quantity $\pi r^2 N$ gives us its total area, when laid out in one plane. From the ratio of these two quantities one obtains the radius of the particle. The result is of the order of a hundred-millionth of a centimeter. This is only one of many ways of estimating this quantity, and the different methods yield surprisingly concordant results, considering how inconceivably small it is.

Molecular dimensions. By following out these lines of argument, and many others (see especially Chap. 39) the values of molecular quantities in Table IX have been obtained for three common gases here given as samples.

Books recommended:

J. Perrin, "Atoms"; translated by D. L. Hammick, 1916 (Constable); especially for Brownian movement.

O. E. Meyer, "The Kinetic Theory of Gases," translated by R. E. Baynes, 1899 (Longmans, Green and Co.).

J. K. Roberts, "Heat and Thermodynamics," (chapter on Kinetic Theory); 1928 (Blackie and Son).

PROBLEMS

1. If all the molecules in a cubic centimeter of hydrogen were laid out in a row just touching one another, how long would this row be? If they were then laid out to cover a surface (assuming them to be square) how large would this surface be?

2. Calculate the average velocity of the particles of sulphur dioxide under standard conditions, given that the density of the gas is 0.0029 grams/cm.^{3}

3. Find the number of gas particles per cm.^{3}, each of mass 10^{-23} grams and moving at an average speed of 1060 m. a second, which would be required to maintain normal atmospheric pressure against the walls of the containing vessel.

4. An automobile tire whose inner volume is 10,000 cm.^{3}, is regarded as being very tight when its gauge pressure falls from 30 to 29 lbs./in.^{2} in a week. Remembering that this pressure is measured in excess of that of the atmosphere, find what fractional part of the pressure is thus lost, and how many air particles must be escaping from the tire in a second. Assume 1 atmosphere = 15 lbs./in.^{2}

5. Following the suggestion of an English physicist, let us suppose that Socrates drank 200 cm.^{3} of poison which was mainly water. How many molecules did this cup contain? (Table IX gives the number of molecules in m grams; m being 18 for water.) If these have since had time to diffuse everywhere, so as to become thoroughly mixed with all the rest of the water on the whole earth, how many of these very same molecules will you drink with your next glass of water (assumed to be 200 cm.^{3})? Take 10^{24} cm.^{3} as the total volume of water on the earth.

6. A perfectly evacuated incandescent lamp bulb of 100 cm.^{3} volume develops a small leak through which a million gas particles rush per second. If they continue to do so indefinitely, how long will it be before the bulb is full?

7. The volume of the earth is 10^{27} cm.^{3} How large would the hydrogen molecules appear to be if 1 cm.^{3} of hydrogen gas were magnified up to the size of the whole earth? How far would they then travel between collisions?

8. The highest "commercial" vacuum is about 10^{-6} mm. of mercury. How many gas particles are there in 1 cm.^{3} of this "vacuum"?

HEAT

CHAPTER 11

TEMPERATURE AND EXPANSION

The nature of temperature. The kinetic theory of gases introduces the reader to the mechanical conception of heat, according to which it is due to the agitation of the particles of matter, whether in a solid, liquid, or gaseous state. According to this idea, the temperature of a body must be connected with the kinetic energy of the particles in their agitation. Turning now to the more purely thermal aspects of this branch of physics, we shall see that this conclusion is in harmony with our idea of temperature as derived from common experiences.

We are familiar with the sensations that bodies produce when they are hot or cold. While these are unreliable except as rough indicators, they furnish us on the whole with a fairly consistent system of temperature ranges, in good agreement with other observations on the effects of heat. We observe that an object which feels hot can always be made to show it in other ways also. For instance, a rod of metal is found to be a little longer when hot than when cold; a quantity of gas when heated exerts a larger pressure than before; a spring gets weaker, etc. Many other examples of this sort could be cited. As a result of all such observations we have adopted a purely arbitrary scale of temperature, based on two fixed points determined by the properties of ice and steam.

Thermometer scales. The first thermometer is supposed to have been invented by Galileo before 1600, and measured differences of temperature only. In 1664 Hooke proposed that the melting

point of ice should be taken as a standard on which to found a scale of temperature. Huygens soon afterwards proposed the boiling point of water for the same purpose, but it was not recognized at the time that *two* points were required to fix a scale which would be practically successful. When this was first suggested, various temperatures were chosen, such as that of a salt-ice-water mixture, the temperature of a deep cellar, the body temperature, the melting-points of ice and of butter, etc. At one time as many as nineteen different thermometric scales were in use. Fahrenheit [1] tried various fixed points, but finally chose the scale we now designate by his name. For this he turned to that most convenient and useful substance, water, to which the founders of the metric system referred when they defined the gram, and from which we shall again derive other units.

The fixed points. The ***lower fixed point*** of the thermometric scale, called 0° centigrade, or 32° Fahrenheit, is the temperature at which ice and water can remain in equilibrium without any change in their relative amounts. It is assumed that the water is pure; small amounts of impurities make an appreciable difference (p. 189). We do not often see a glass containing ice and water which continues to keep them in the same proportions; usually the ice diminishes and the water gains thereby. This, however, is because the glass is surrounded by warmer bodies; if it were kept in a region at its own temperature, there would be no change in the proportion of the two constituents.

The ***upper fixed point,*** or "steam point," called 100° centigrade, or 212° Fahrenheit, is defined as the temperature existing in steam above boiling water, under standard atmospheric pressure (76 cm. of mercury). Once more, the water should be pure, though here impurities do not affect the temperature unless the water spatters over the bulb of the thermometer (which it is very likely to do). The pressure must be mentioned in this case because its effect is large (p. 194), but in the case of the lower fixed point it can be neglected.

Centigrade and Fahrenheit scales. The centigrade scale of temperature contains 100 degrees between the fixed points. This is the scale used in many European countries, and by scientific men the world over. The Fahrenheit scale has 180 degrees be-

[1] Gabriel Fahrenheit, 1686–1736, physicist and instrument-maker; born in Germany, though long a resident of Holland and England.

tween the same points. The zero of the latter was planned by its maker to be as low a temperature as could be reached artificially. It is an interesting comment on scientific progress to note that the technique of low temperature work has so improved since Fahrenheit's day that a temperature 458 of his degrees below zero has been attained. The Fahrenheit scale has nothing to recommend it, but we cannot ignore it since it is in domestic use in most English-speaking countries. Its degree is five-ninths as great as the centigrade one; and the scale starts from a different zero. To convert from one scale to the other it is necessary to

TABLE X

	Centigrade	Fahrenheit
The interior of hot stars, up to	30,000,000	54,000,000
Their surfaces	20,000	36,000
The sun's surface	5,700	10,000
Carbon arc in air, up to	3,500	6,330
Tungsten lamp, gas-filled	2,900	5,200
Bunsen flame about	1,750	3,180
Coal furnace, up to	1,500	2,700
Dull red heat	600	1,100
Boiling point of mercury	356.7	674.1
Boiling point of water	100	212
Blood heat	37.0	98.6
Common salt and ice mixture	− 20.	− 4.
Mercury freezes	− 40.	− 40.
Liquid oxygen	− 183.	− 297.
Liquid helium	− 268.6	− 451.5
Lowest attained	− 272.36	− 458.25

note that $t_F - 32$ is the number of Fahrenheit degrees above 0° C. Therefore $\frac{5}{9}(t_F - 32)$ is the number of centigrade degrees above 0° C., or is t_C. From which $t_F = \frac{9}{5}t_C + 32$. This formula enables one to make the necessary conversions easily. Anyone who wishes to go far in scientific work will find it a great convenience to become familiar with the centigrade scale.

To assist in this process, and for its own interest, the range of common temperatures is illustrated by the data on the two scales in Table X.

The mercury thermometer. A thermometer is any type of instrument for the measurement of temperature. All sorts depend on heat effects, (changes in dimensions, pressure, electrical resistance, heat radiated, etc.) but these are so various that the important

types of thermometers will not all have been mentioned until we have considered the effects of temperature on electrical and optical properties. For the present, only one, the commonest type of thermometer will be described. Others will be found on pages 180, 194, 339, 367 and 534.

A thick-walled glass tube with a very fine central hole down its axis is drawn as uniformly as possible, and sealed to a thin-walled bulb, usually cylindrical (Fig. 11-1). This is then filled with mercury up to some point in the tube, which is done by first inverting it in mercury when warm, and allowing it to cool, thus drawing a little mercury into the bulb. This process is repeated, and the last air is driven out by boiling the mercury. The top of the tube is sealed off while the tube contains nothing but mercury and its vapor. The positions of the mercury surface at 0° and 100° C. are marked on the tube, and the space between is graduated. The scale may be extended in either direction beyond the fixed points. This form of thermometer has a range sufficient for ordinary experiments, but has certain defects. One which may sometimes be observed is a lowering of the ice-point after the thermometer has been heated. This makes all readings depend a little on the previous history of the instrument. The effect is explained as a sort of elastic lag, (p. 126) such as can be produced in ordinary rubber bands. The glass bulb of the thermometer is expanded by heating, and does not immediately return to its original volume. Special "thermometer glass" has been made which is practically free from this effect, and the better instruments are made of it.

Fig. 11-1 Mercury thermometer

Expansion. The action of the mercury thermometer depends on expansion by heat; in this case on the difference of expansion of mercury and glass. Solids, liquids, and gases all show this effect. When a substance is being heated the agitation of its particles pushes them a little farther apart, on account of the increased vigor of their impacts. Solids show but a small change; some in fact practically none. A few curious alloys even shrink with rise of temperature. Evidently no simple explanation of thermal expansion is adequate for all cases.

Linear expansion of solids. A common way of measuring the expansion of solids is to heat a long bar, or tube, of the solid with

TABLE XI

Linear Coefficients of Expansion for Solids

(Expansion of 1 cm. for 1° C. change)

Iron	11×10^{-6}	Mahogany along grain	3×10^{-6}
Brass	18 × "	" across "	40 × "
Aluminum	25 × "	Pine along grain	5 × "
Platinum	8.9 × "	" across "	34 × "
Invar	0.9 × "	Glass (soda)	8.5 × "
Fused quartz	.0.42 × "	" (pyrex)	3. × "

steam, or in a small furnace, and observe the small change in length by means of a microscope, multiplying lever, or micrometer. Experiments of this sort are commonly performed in the laboratory; by refined methods exceedingly minute changes are easily measurable. The amount of the effect observed depends on the kind of material, the change of temperature, and the length of the specimen. Each centimeter of length expands for one degree change of temperature by an amount characteristic of the material and known as its ***linear coefficient of expansion.*** If c is this coefficient, L the length of the specimen, t the change of temperature, and e the change of length, it follows that

$$e = Lct.$$

Strictly speaking, the rate of expansion depends somewhat on the particular range of temperature selected. It is usual to begin at 0° C., with a length L_0. Then the length at any other temperature t is found from $L = L_0(1 + ct)$. Here t is both the temperature and the change of temperature.

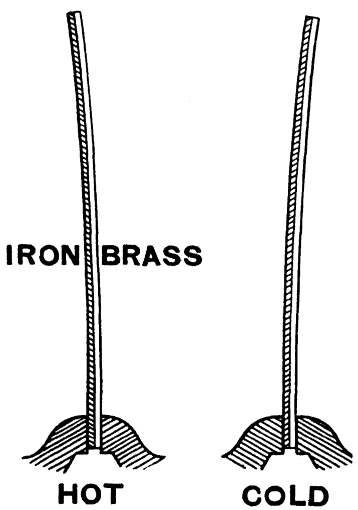


Fig. 11–2

Applications of expansion. If long, thin strips of different materials, such as iron and brass, are fastened firmly together, face to face along their length, and if one end of this compound strip is held fixed, the free end will move considerably for small changes of temperature (Fig. 11–2). Such devices are used with multiplying attachments as metallic thermometers; or their motions may be made to close electric circuits, and thus

control the fires in furnaces, as in domestic thermostats. If such a strip is made in the form of a semicircle, the diameter of the circle will change with temperature. In the balance wheel of a watch, this device is used to diminish the rotational inertia with rise of temperature, and so compensate for the weakening of the elastic force of the hairspring due to the same cause (Fig. 11–3). Thus the rate of oscillation may be made independent of the temperature.

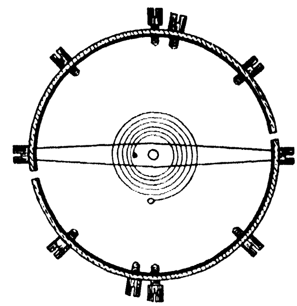


FIG. 11–3

Volume expansion. The change in volume which a solid undergoes must often be found. If a cube has each side of length L its volume will increase according to the formula

$$L_t^3 = L_0^3(1 + ct)^3,$$

L_t being the length at $t°$ C.; but, since c is very small we may neglect terms involving c^2 and c^3; hence on expanding the right-hand side we get

$$L_t^3 = L_0^3(1 + 3ct).$$

If the body is not in the shape of a cube the same relation holds; that is, its volume V is given by

$$V = V_0(1 + 3ct).$$

Evidently the ***volume coefficient of expansion*** is approximately $3c$; that is, three times the linear coefficient.

Further applications. The apparently uninteresting expansion coefficients are sometimes of great practical importance. The successful manufacture of incandescent lamps was first accomplished by using platinum wires to conduct the electric current through the glass wall of the bulb. This was because platinum was the only metal having nearly the same expansion coefficient as the sort of glass used. Hence, if sealed into the glass when hot, it did not strain and crack the seal when cool. The high cost of platinum made it necessary to find a cheaper material with the same coefficient, and this was done by making an alloy of iron and nickel, which is now used as a "platinum substitute." Accurate surveying is possible only by taking into account the change in length of the measuring tapes or chains with temperature; but since the discovery of the nickel-steel alloy "invar" this change has been practically eliminated by making the tapes of this material. Allowance for expansion is sometimes made in steel bridges by supporting one end on rollers. The amounts to allow are readily calculated beforehand.

Force of expansion. If a substance is to expand, a large force must act to compel the cohesion of its particles to yield a little. The kinetic agitation of the particles furnishes this force. If one attempts to prevent the expansion, he finds that the forces involved are extremely large. Forces of the same order of magnitude must arise in compressing a substance as in preventing it from expanding when the temperature rises. Hence, if one substance expands more than others, this is an indication that its molecular forces are weaker, and it should be more compressible. A relationship of this sort between compression and expansion coefficients has actually been found to hold in many cases, though not universally.

An experiment can be arranged to show the magnitude of the forces of expansion. If a very strong frame is provided, Fig. 11–4, a bar can be placed in it so that one end pushes against the middle of a short steel rod, which will be broken by the force of expansion when the rod is heated.

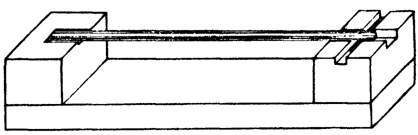


Fig. 11–4

When steel rails are welded together, as is often done in city trolley systems, no allowance can be made for expansion. If the rails are so laid as to be unable to escape, they will compress each other longitudinally on a hot day, the compression rising to such a value as to balance the expansive force. The steel is elastic and returns to its former condition when cooled again.

The expansion of mercury. The expansion of liquids is usually larger than that of solids and varies more with the range of temperature considered. Two cases are of especial interest. The expansion of mercury is important because of its use in thermometers. While mercury is the best available liquid for this purpose, it is not perfect. Its rate of expansion increases slightly with the temperature, with the result that from 0° to 50° C. its expansion is less than from 50° to 100° C. Hence a mercury thermometer with uniform bore and uniform graduations, with the 50° mark exactly midway between 0° and 100° will read low when the temperature is exactly 50° and the error amounts to one or two tenths of a degree. Naturally, this error is not detectable unless one first obtains a thermometer free from it. Such a one is the gas thermometer, considered below (p. 157).

The expansion of water. The expansion of water is unusual. Between 0° C. and 4° C. it shrinks instead of expanding, and its rate of expansion continually increases from 4° to 100° C. The volume of a quantity of water carried through the range 0° to 10° C. is shown in Fig. 11–5. The rate of expansion will be given by the slope of this volume curve, and this is plotted in Fig. 11–6 against the temperature, showing its steady growth from a negative to a high positive value. This anomalous property of water is useful. A pond in cooling below 4° C. keeps its coolest water on top, and freezes first there. Fortunately, also, ice is lighter than water, so that the ice remains on top, forming a protecting coating tending to retard the process, and thus making it easier for aquatic life to survive below.

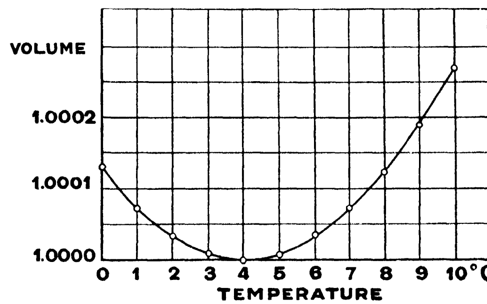


FIG. 11–5

The changes in volume of watei from 0° C. to 10° C.

In the case of salt water, it is worth noting that the maximum density occurs at about 0° C.

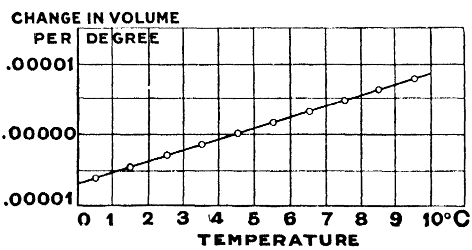


FIG. 11–6

The rate of expansion of water from 0° C. to 10° C.

Measurement of Liquid Expansion. The expansion of liquids as usually examined is combined with that of the containing vessel. Thus water in a glass vessel will appear to shrink in volume between 4° C. and 5° C. In reality it expands a little, but the glass expands faster. The apparent expansion of a liquid, then, is really the difference between its expansion and that of the vessel containing it. To avoid this complication and obtain the volume expansion of the liquids directly, Dulong and Petit[1] de-

[1] Pierre Dulong (1785–1838), French chemist and physicist, who carried out important experimental researches on Boyle's law, the temperature of high-pressure steam, specific heat, and many other topics in heat and light. In several of these he was associated with Alexis Petit (1791–1820), whose successor he became as Professor of Physics at the École Polytechnique in Paris.

vised an interesting experiment. Two vertical glass tubes connected together, as shown in Fig. 11–7, are filled with the liquid to be tested. One of them may be surrounded with cracked ice, and the other enclosed in a jacket filled with steam. If the two tubes are at different temperatures, they virtually contain different liquids (as on p. 24) balanced against each other, exerting the same pressure at the bottom B. This pressure is numerically equal to hd, or to $h'd'$, where h and h' are the heights and d and d' the corresponding densities. Hence

$$\frac{h'}{h} = \frac{d}{d'} \quad \text{and} \quad \frac{h' - h}{h} = \frac{d - d'}{d'}.$$

This relation is true no matter what may be the shape or size of the tubes themselves. Now if we consider any mass m of a liquid of density d and volume v, and let it expand to volume v' and density d', we shall have

$$m = dv = d'v'.$$

Hence

$$\frac{d}{d'} = \frac{v'}{v} \quad \text{and} \quad \frac{d - d'}{d'} = \frac{v' - v}{v}$$

which is the change in volume per unit volume. The volume coefficient of expansion, k, is the change in volume per cm.³ per degree; hence

$$k = \frac{v' - v}{vt} = \frac{d - d'}{d't} = \frac{h' - h}{ht}$$

where t is the change of temperature. This last value is easily measurable, and yields us the coefficient of expansion of the liquid *without considering that of the glass.*

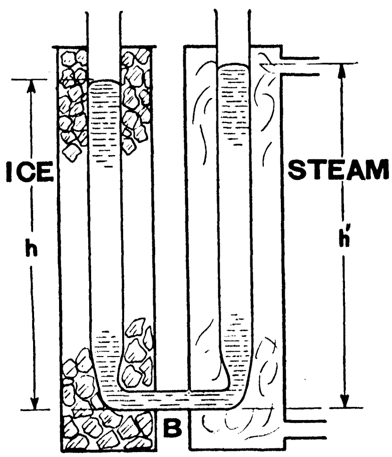


FIG. 11–7
Apparatus for measuring the expansion of a liquid

The expansion of gases. The study of gas expansion yields results of great importance. There are two types of experiment through which we might examine it. Both can be performed with an apparatus known as the gas thermometer. In one method the gas is allowed to expand at constant pressure, changes in volume being noted; in the other the volume is kept constant while changes in pressure are determined. The gas is enclosed in the bulb B (Fig. 11–8), which may be surrounded by ice, steam, or water in an appropriate jacket. The bulb is connected to the tube G by a bent tube C of small bore and to a second tube H through a flexible tube R. The tubing GRH is filled with mercury.

The constant volume gas thermometer. If this apparatus is used at constant volume, a mark M is made on the glass, and by raising or lowering the tube RH the surface of the mercury is always brought to this fixed mark. This is done first when the bulb B is surrounded by cracked ice, and the position of H is then marked 0° on the scale; then B is put in a steam bath, which makes the surface in G tend to go down. The tube RH has to be raised to keep it at M. The new position of H is then marked as 100°. If B is kept at any intermediate temperature, the height of H on the scale thus gives a reading of this temperature and the instrument acts as a thermometer. It proves to be a singularly regular and accurate one. In the better forms of the instrument the tube C is narrow enough to be of negligible volume, or for the most accurate work there are means of compensating for the slight error introduced by the fact that the gas in C is not all at the temperature of B.

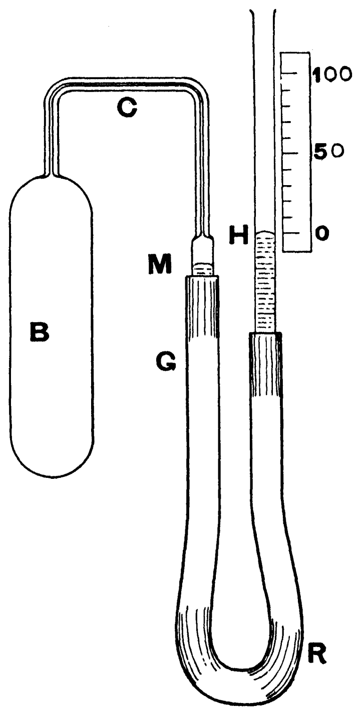


FIG. 11–8
A gas thermometer

Measurement of the absolute zero. In the arrangement just considered, what is measured is the change of pressure in the gas. The tube H is open to the atmosphere, and the total pressure on the gas is equal to that on the surface H plus that due to a mercury column whose vertical height is MH. If the barometric height is added to MH, the total pressure P on the gas in B is obtained. If the values of this pressure are plotted as in Fig. 11–9 at the two fixed points 0° and 100° C., and these points E and F when carefully determined are joined by a line, this line may be considerably prolonged in both directions, and will then faithfully represent the behavior of such gases as hydrogen or helium. Its prolongation to the left leads to a point where the pressure is zero, which is at about $-$ 273°.18 C. This is the temperature of the absolute zero, at which the particles of the gas

TABLE XII

Expansion Coefficients of Gases (0° *to* 100° *C.*)

	a (constant pressure)	b (constant volume)
Hydrogen	0.0036610	0.0036625
Nitrogen	3673	3672
Oxygen	3668	3674
Helium	——	36625
Carbon dioxide	3707	3698
Sulphur dioxide	3903	3845

would presumably be at rest. No lower temperature, according to the kinetic theory, is conceivable. Though actual gases liquefy, and even solidify, before reaching that temperature, the gas helium remains usable in a gas thermometer to within a few degrees of its boiling point (−268°.6 C.). There are other lines of argument (p. 144) which point to this same temperature as the lowest possible. In 1920 Kamerlingh Onnes[1] produced and measured a temperature within 0°.82 of that point, by methods that will be discussed later (p. 205).

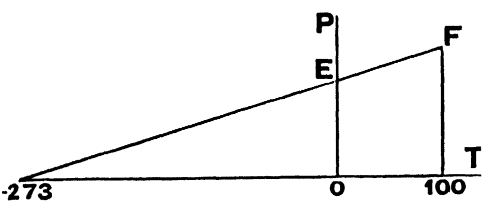


FIG. 11–9
The changes of gas pressure with temperature

Absolute temperatures. If we make the absolute zero the starting point of our temperature scale, we get a more rational arrangement; any scale of this sort is called a ***scale of absolute temperatures.*** The absolute scale commonly used is the one whose degree is the same as on the centigrade scale; we use the letter T for absolute temperatures, reserving t for degrees centigrade. The custom is to write "$T°$ abs.," or more often "$T°$ K."; the latter abbreviation refers to the Kelvin scale (p. 222). On the absolute scale, the diagram in Fig. 11–9 becomes a little simpler. The straight line EF then passes through the origin, which means that the pressure in the gas is proportional to the absolute temperature; or, $P = bT$, where b is the constant of proportionality, or the change of pressure per degree. Its value is 1/273, (or about 0.00366) as the diagram shows. Careful experiments have yielded the values shown in Table XII. It will

[1] Late professor in the University at Leyden, Holland.

be seen that the constant b is nearly the same for all the lighter permanent gases.

Gas thermometer at constant pressure. The gas thermometer may also be used to measure the rate of expansion of a gas at constant pressure. The surfaces G and H (Fig. 11–8) are to be kept always at the same level, when the temperature changes; then the pressure of the gas in B will always be equal to that of the atmosphere. (If this is changing, allowance must be made for it.) At low temperatures G and H will be high, at high temperatures low. The scale will therefore be drawn with 0° C. above; but otherwise the instrument is used as before, and the scale proves to be of the same size. If the volume of B is known, the actual changes in volume can be measured and the rate of volume expansion determined. A curve like that in Fig. 11–9 can be drawn, which proves in fact to be identical with it, except that V (volume) takes the place of P. It follows that $V = aT$; the coefficient of expansion a for different gases is given in Table XII and will be seen to be very nearly the same as the coefficients of pressure change, b, already considered.

Ranges of thermometers. The gas thermometer can be used over a very wide range of temperature. Its lower limit is reached when helium gas is used in it and this gas approaches its boiling point; its upper limit, when the walls of the bulb (best made of fused quartz) begin to soften or become porous to the gas within, at about 2000° C. An ordinary glass bulb softens near 500° C. A mercury thermometer is usually limited in range by the freezing (− 40° C.) and boiling (357° C.) of the mercury; though when the space above the mercury is filled with nitrogen or carbon dioxide gas, the boiling may be prevented and the range raised to 500° C., or even higher with especially hard glass. Thermometers made of fused quartz may be filled with the metal gallium and used at still higher temperatures. For temperatures as low as those of liquid air, a pentane-filled glass bulb thermometer can be used. No liquid thermometer can be made without being calibrated in comparison with a gas thermometer, and even then other types are to be preferred for accuracy. For temperatures above 2000° C. quite different methods of measuring temperatures are available.

The gas laws. We have seen that if the volume of a gas is kept constant, it follows that P is proportional to T; also, if the pressure is kept constant, V is proportional to T. If the temperature is kept constant, PV is constant (Boyle's law). These three statements can be combined into one, which is that PV is proportional to the absolute temperature T. This statement

holds for any given mass of gas, but the constant varies with that mass. The law is usually stated as $PV = RT$, and in this case it is understood that the mass of gas chosen is that known to chemists as the gram-molecular weight; that is, a mass whose weight in grams is equal to the molecular weight of the gas. The volume of this mass under standard conditions is the same for all gases, (equal to 22,415 cubic centimeters.)[1] Since $R = PV/T$, its value is readily calculated; if P is in dynes per square centimeter, V in cubic centimeters, and T in degrees absolute, the value of R becomes 83.15×10^6. Since PV is a quantity of work or energy (p. 221) it is thus expressed in ergs, and R in ergs per degree per gram-molecule. If desired, R may be expressed also as 1.987 calories per degree per gram-molecule using the heat unit shortly to be defined.

The general gas law is easily used in the solution of problems. If it is written in the form

$$\frac{P_1V_1}{T_1} = \frac{P_2V_2}{T_2},$$

where the subscripts refer to different states of the same mass of gas, and five of these six quantities are known, the sixth is easily found. Sometimes changes in density are desired, in which case the fact that density (D) and volume vary inversely enables us to obtain the relation

$$\frac{P_1}{T_1D_1} = \frac{P_2}{T_2D_2}.$$

Deviations from the gas laws. Unfortunately, ordinary gases do not follow these gas laws quite accurately, though the deviations are often very small. When Boyle's law is tested with high precision, it is found that no gas follows it *exactly* over more than a short range. In the case of all gases except hydrogen at ordinary temperatures, the product PV decreases slightly at first with increase of pressure, especially at low temperatures. Thus, with air the decrease at 10 atmospheres of pressure is about 1%, while in the case of carbon dioxide the decrease in the same range is 8%. For hydrogen on the contrary there is a slight increase. At some particular temperature each gas follows the law exactly over a short range, and all act like hydrogen at tem-

[1] This follows from the following reasoning: Avogadro's rule (p. 144) states that each cubic centimeter of gas contains the same number of molecules; hence 1 gram of gas must occupy a volume inversely proportional to the molecular weight; or a number of grams equal to the molecular weight must occupy a constant volume.

peratures high above their boiling points. At very high pressures, such as Bridgman has investigated (up to 20,000 atmospheres) the product PV increases for all gases. Figure 11–10 shows one of his curves for nitrogen.

This seems at first sight rather a blow to our kinetic theory; but it shows merely that we have neglected one or two effects which are small at low pressures. Since we found attraction (cohesion) to exist between the particles in solids and liquids, it seems natural to suggest that there may be a trace of this

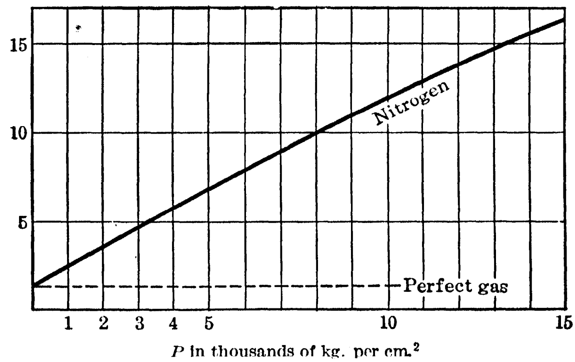


FIG. 11–10

Values of PV for nitrogen at high pressures

effect in gases also. Moreover, in the elementary deduction of Boyle's law (p. 143) no account was taken of the fact that the particles themselves occupy a portion of the volume in which we supposed they had perfect freedom to move about. If a diminution of the free volume by the amount of the volume of the particles themselves is admitted, as well as a slight attraction between them, we arrive at Van der Waals' modification of the laws of gases, in the form

$$\left(P + \frac{A}{V^2}\right)(V - B) = RT.$$

Here $\frac{A}{V^2}$ is an added pressure due to the molecular attractions, and B is the volume due to the particles themselves. This formula is a great improvement over the "ideal gas" formula, but real gases deviate even from it when more precise determinations are made over the widest ranges of temperature and pressure. It is, however, easy to use, and of sufficient precision for many purposes.

PROBLEMS

1. Make a graph showing the relation between the centigrade and Fahrenheit scales.

2. Find the temperature at which the centigrade and Fahrenheit scales agree.

3. Find the centigrade temperatures corresponding to 0° F. and 70° F.; and the Fahrenheit temperatures corresponding to 600° C., and − 10° C.

4. Lead sheets when laid on sloping roofs are apt to creep down the slope unless very well fastened, on account of the oscillations of temperature to which they are subjected. Explain.

5. A brass meter bar is found to be longer than an iron one at 25° C. by 0.007 cm. At what temperature do they agree?

6. A distance is to be measured by means of a steel tape graduated in feet (and decimals), whose coefficient of expansion is 0.000011. The tape is correct when it is at 15° C., but the measurement is made on a hot day when it is 25° C. If the length measured was 1000.53 ft., what was the true length?

7. A block of iron has a spherical cavity inside it whose volume is just 1 cc. at 0° C. What is the volume of the cavity at 100° C., if the coefficient of linear expansion of the iron is 0.000012?

8. The linear coefficient of a sort of glass is 0.000009. What would the density of a piece of this glass become at 20° C., if its value were 2.5 at 0° C.?

9. An iron ball, 10 cm. diameter, is 0.01 cm. too large to go through a hole in a brass plate at 0° C. At what temperature will it just pass through, if the ball and plate change temperature together?

10. The coefficient of expansion of mercury is 0.00018, and its density at 0° C. is 13.6. What is its density at 100° C.?

11. A tall narrow glass vessel, graduated in cubic centimeters, contains 40 cm.^{3} of lead shot, and enough oil to fill it up to the 100 cm.^{3} mark, all at 0° C. If the temperature rises to 50° C., at what point will the surface of the oil stand? The volume coefficient of the glass is 0.000027; of lead, 0.000057; of oil, 0.0004.

12. A vertical glass tube of 1 mm.^{2} internal cross-section is closed at the bottom and filled to a height of 1 m. at a temperature of 0° C. If the linear coefficient of expansion of glass is 0.000009, and the volume coefficient of mercury is 0.00018, find the height of the mercury in the tube at 50° C.

13. Find the increase in volume of a glass vessel whose volume at 0° C. is 100 cm.^{3}, when it is raised to 60° C. If this vessel when at 0° C. had been filled with mercury, how much mercury would have spilled out during the rise of temperature? Take the linear coefficient for glass as 0.000009 and the volume coefficient for mercury as 0.00018.

14. If the pressure in a large tank containing 1000 cu. m. of illuminating gas remains constant, how much does the volume change when the temperature falls from 27° C. to 17° C.?

15. By what fraction of its volume is the air entering a hot-air furnace at − 10° C. increased when it emerges at 50° C.?

16. If the volume of 1 gram of air at 0° C. and 76 cm. pressure is 773 cm.^{3}, what is the weight of air in a cubic meter at 15° C. and 74 cm. pressure?

17. A bottle containing dry air is closed air-tight when out-of-doors in winter at a temperature of − 20° C., and a pressure of 76 cm. Find the

pressure in the bottle when it is brought indoors and put into boiling water (100° C.). Show why the expansion of the bottle may be neglected.

18. An automobile tire, whose volume may be supposed to be constant, is pumped up to a pressure of 40 lbs. per square inch (as indicated by the gauge) at a temperature of $-10°$ C. The car is then run into a warm garage where the temperature is 20° C. Find the pressure in the tire then, assuming no leakage. (N.B. The gauge registers pressure *above that of the atmosphere.*)

19. A rubber balloon whose volume is 100 cu. m. rises from a level where the pressure is 75 cm. and the temperature $+ 7°$ C. to a height where the pressure is 25 cm. and the temperature $- 23°$ C. Find its volume, assuming it to expand freely.

20. A hot-air balloon whose volume is 20 $m.^3$ is filled with air at 40° C. The outer air is at 15° C. What is the total lifting power of this balloon?

21. The density of ordinary hydrogen is 0.000089 grams per $cm.^3$ A steel cylinder of one-fifth of a cubic meter capacity is filled with hydrogen at a gauge-pressure of 20 atmospheres. How much does this hydrogen weigh?

CHAPTER 12

QUANTITY OF HEAT

Units of heat, 164; specific heat, 164; atomic heat, 165; the mechanical equivalent of heat, or the heat equivalent of work, 166; heating by compression, 167; the Joule-Thomson effect, 168; the total heat in a body, 169; specific heat of gases, 169; the method of mixtures, 170; the continuous-flow calorimeter, 171; the accuracy of heat experiments, 172; experiment on specific heats, 172.

Units of heat. In the early days of physics, heat was thought of as a weightless fluid, which could be poured into a body, or emptied out of it, and it was natural to conceive of this as having bulk or quantity. Now that this idea has been discarded, we retain a few remnants of the same terminology, and still speak of the quantity of heat put into a body. The unit of this quantity is obtained by referring once again to the properties of water. The quantity of heat required to raise one gram of water through one degree centigrade is called a ***calorie***,[1] sometimes a small calorie or gram-calorie. Those who deal with experiments on a large scale prefer what is called the large calorie, or the kilogram-calorie, which is a thousand times as great. This unit is used, for instance, in considering the food requirements of the human body. The corresponding English unit is the British thermal unit (B. t. u.) which is the heat required to raise one pound of water through one degree Fahrenheit; it is equal to 252 calories.

Specific heat. Water is a remarkable substance. It requires more heat to warm it, and gives off more heat in cooling through a given range of temperature than any other liquid, or than any solid, with the exception of only a few rare materials (e.g. liquid lithium). As important practical problems involve heating many other substances beside water, we need a name for the amount of heat required to warm a gram of any substance through one

[1] In work of high precision the particular degree between 15° and 16° C. is sometimes specified; or that between 17° and 18°. Neglect of this refinement introduces errors of less than one per cent, which, considering the disappointingly low accuracy usually obtained in heat measurements, is unimportant here.

degree. We define the ***specific heat*** of a substance as the ***ratio*** of the amount of heat required to warm one gram of the substance through one degree centigrade to the amount similarly required for water. It is numerically equal to the number of calories needed to warm one gram of the substance through 1° C., but its physical "dimensions" (p. 76) are different. The specific heat, being a ratio, is a pure number with no dimensions; and it has the same value in either system of units.

The specific heats of a number of substances are given in Table XIII for the temperatures indicated.

TABLE XIII

Specific Heats

Aluminum	− 240° C.	0.0092	Lead	0° C.	0.030
	0°	0.2096		300°	0.034
	600°	0.282	Brass	0°	0.09
Copper	− 250°	0.0035	Glass	0°	0.16
	0°	0.091	Ice	0°	0.50
	100°	0.095	Alcohol	0°	0.55
Iron	0°	0.105	Mercury	0°	0.0335

The relatively low values of the specific heats of heavy substances are to be noted. The specific heats of most substances increase appreciably, sometimes very rapidly, with the average temperature. That of water changes in the manner shown in Fig. 12–1.

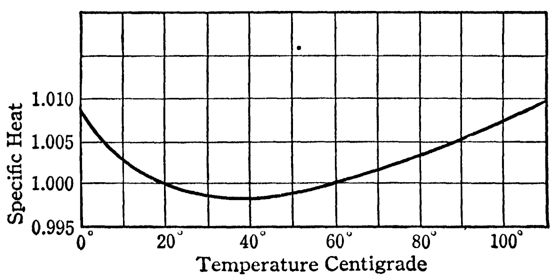


FIG. 12–1

Changes in the specific heat of water with the temperature

Atomic heat. Dulong and Petit (1819) discovered an interesting fact which follows from the values in Table XIII. A relation holds for a large number of chemical elements in solid form and for many other materials, namely, that the product of the specific heat by the atomic weight is nearly constant at ordinary tem-

peratures (the value is about 6.3). If we call the heat needed to warm one atom one degree the ***atomic heat,*** we see that

$$
\begin{aligned}
\text{specific heat} \times \text{atomic weight} & \\
&= \text{heat to warm 1 gram } 1^\circ\text{ C.} \times \text{atomic weight} \\
&= \text{atomic heat} \times \text{number of atoms in 1 gram} \times \text{weight of each} \\
&= \text{atomic heat}
\end{aligned}
$$

since the number of atoms times the weight of each is just 1 gram. Thus the law found by Dulong and Petit indicates that the atomic heat is nearly constant. It is not true of all the solid elements (e.g. silicon, carbon, etc.), but holds at ordinary temperatures with fair accuracy for most of the metals, while at high temperatures it becomes more generally true.

The mechanical equivalent of heat, or the heat equivalent of work. The modern conception of the nature of heat dates back to experiments by Rumford[1] (1798), Joule (1849), and others, showing that when mechanical work disappears a definite equivalent quantity of heat reappears, no matter how the method is varied. The work may be produced by ordinary friction, by viscosity in liquids, by compression of gases, or by magnetic and electric means, always with the same result. This is precisely what we should expect from the kinetic theory. If work is done on a gas by compressing it, its temperature must rise (p. 167); if work is done by friction, the energy must go into a kinetic agitation of the particles which are involved in the action. A good experiment on this sort of effect may be performed, after Joule, by rotating a paddle wheel in a vessel (Fig. 12–2) supplied with fixed vanes projecting inward from the sides. The water in the vessel becomes a mass of eddies when the wheel rotates. The force needed to turn such a device is much larger than our experience in stirring ordinary liquids would suggest, due to the almost complete absence of steady flow in the vessel. Falling weights can be arranged to turn the wheel so that the work put

[1] Born as Benjamin Thompson, (1753–1814) in Massachusetts; he spent a picturesque life as statesman, military expert, and scientist, largely in England and Bavaria. He was made a Count of the Holy Roman Empire in 1791, and chose the name of his wife's old home, Rumford (now Concord), N. H. He was largely responsible for originating the scientific laboratory known as the Royal Institution in London, and left various funds for scientific endowments. Beside his work on heat, he invented a photometer, and applied his originality to solving many problems in domestic (as well as military) science.

into the water can be measured; or, by using an electric motor to run it, as Rowland did, the work may be measured electrically (p. 372) and the operation may be prolonged indefinitely. The slow rise of temperature of the water can be noted, but, as usual, precautions must be taken to prevent the temperature from being affected by loss or gain of heat due to other causes. From such an experiment, Joule obtained an accurate value for the ***number of work units needed to provide one heat unit,*** a quantity which is usually called the ***mechanical equivalent of heat,*** though it is more logical to name it the ***heat equivalent of work.***[1] It is variously stated as 41,800,000 ergs per calorie, 4.18 joules per calorie, 427 gram-meters per calorie, or 778 ft.-lbs. per B. t. u.

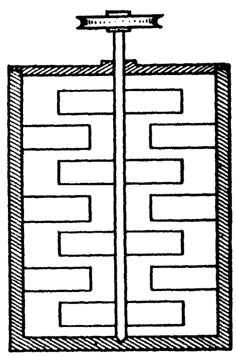


FIG. 12–2

Heating by compression. Another example of the production of heat by work is furnished when a gas is heated by being compressed. This experiment is one that may be used to measure the heat equivalent of work. By a sudden compression exerted on a small quantity of air in a strong cylinder the temperature may be raised so much that a bit of tinder inside may be ignited. This is known as the experiment of the "fire syringe." A violent compression in a gasoline engine cylinder may, especially in the presence of finely divided carbon on the walls, produce such a rise in temperature as to cause an explosion, usually before it is wanted; this is then known as pre-ignition. In the Diesel engine, heating by compression is relied upon to ignite the explosive mixture.

The kinetic theory explains this rise of temperature very simply. A baseball which is struck by a heavy forward-moving bat rebounds from the bat with a greater speed than it had on the approach. Similarly, a gas particle striking a piston which is moving to meet it, rebounds with a higher speed; therefore the gas particles after such treatment have greater kinetic energy than before, and this implies a higher temperature. The converse is also true; the gas particles move more slowly after colliding with a

[1] Actually, the mechanical equivalent of heat is what we try to get out of engines, and the amount we get depends on the conditions under which the engine works (see Chapter 15).

receding piston, hence, a gas is cooled by an increase in volume. In this case the gas particles, which exert a pressure against the piston, ***do work upon it,*** and hence lose energy, and fall in temperature. A compressed gas, escaping into the open air, cools itself in like manner, because it has to push aside the atmosphere in order to make room for itself, and hence, it does work. One has only to put a finger in the stream issuing from an open valve in an automobile tire to feel the coolness of the escaping air.

The Joule-Thomson effect. If an experiment were arranged in which a compressed gas had a chance to expand into an ***empty*** space, our simple kinetic theory indicates that no work would be done, and no cooling should occur. This can be arranged by having two bulbs, one exhausted, the other full of gas, with a connection between which is closed by a cock (Fig. 12–3). Since the particles of the gas do not all possess the same speed, the faster particles will be more likely than the slow ones to enter the exhausted vessel when the cock is first opened; so that the gas in that vessel will be slightly warmer at first than the gas left behind in the first vessel; but unless there are forces between the gas particles themselves, which we did not postulate when we began to look into the kinetic theory of gases (p. 139), there will be no work done, and the gas ***on the whole*** will neither gain nor lose heat. This difficult experiment was tried by Joule [1] without success because the effect sought for was very small and the experimental error rather large. Later (1854) Joule and Thomson [2] repeated it in a modified form which is known as the ***porous plug experiment*** and obtained positive results. In this form a gas kept at

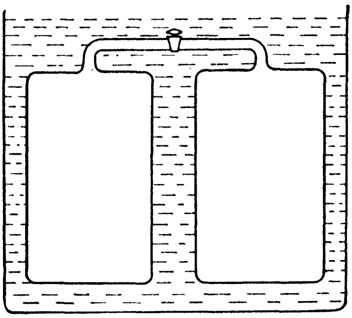


Fig. 12–3
The Joule-Thomson experiment (much simplified)

[1] J. P. Joule (1818–1889); an English brewer who took physics as his hobby, and did very important research in heat and electricity. A unit of work is named after him, as well as the law showing how much heat is developed by an electric current in a circuit.

[2] William Thomson (1824–1907), became Lord Kelvin in 1892; probably the greatest English physicist of the nineteenth century; made the first oceanic cable possible, developed many instruments for navigation and carried out many fundamental researches, especially in heat, electricity and light.

constant pressure by a pump is allowed to flow continuously through a porous plug in a pipe into a region kept at a much lower pressure. The resistance offered by the plug to the flow is so large that the gas can be treated as at rest on both sides of the plug; that is, the kinetic energy of the flow is negligible. The gas merely expands through the plug. The action is steady, and thermometers placed in the gas at each side of the plug give its temperature on entering and on leaving. The ***lowering of temperature*** which was actually found in all gases except hydrogen can be explained only by assuming that there is a very slight attraction between the gas particles which has to be overcome when they are separated. The work used up in this way is taken from the kinetic energy of the gas particles, and cooling results. We have already seen reason to assume some such attraction (p. 161) on other grounds.

The ***Joule-Thomson effect,*** as this phenomenon is called, becomes quite large at very low temperatures, and all gases then behave alike.

The total heat in a body. In the case of a gas which is heated at constant volume, a large part of the energy put into it takes the form of increased kinetic energy of its molecules, either as erratic motions of the sort already described, or as vibrations of the parts of the molecules themselves. This is not so true of solids and liquids; with them some of it takes the form of potential energy also. It has already been noted (p. 124) that large forces are involved in compressing or stretching solids. These are necessary to overcome the cohesive forces. The expansion which occurs on heating must also be accomplished in spite of the opposition of these same large forces, and in this operation a quantity of work will be stored away in the form of potential energy in the expanded substance. Hence the temperature, which depends on the kinetic energy, is no longer a measure of the total energy in the material, and equal masses of different bodies at the same temperature contain different amounts of heat. A cup of warm water may contain much more heat than a lump of red-hot iron of the same weight. The distinction between high temperature and high heat-content is an important one to grasp, and is closely connected with the idea of specific heat.

Specific heat of gases. A gas can be heated under two quite different conditions. It may be kept at constant volume, in which case all the energy used in heating it goes into increased agitation

TABLE XIV

Specific Heats of Gases

(in calories per gram)

	At constant pressure	At constant volume
Dry air	0.242	0.172
Hydrogen	3.40	2.40
Carbon dioxide	0.20	0.165
Steam (100° C.)	0.48	0.34

of its particles; or it may be kept at constant pressure in some sort of elastic container which will allow it to expand. In the latter case it does work in expanding, pushing the atmosphere out of the way, and thus it must be supplied with more energy in order to reach the same temperature. Hence it follows that a gas has ***two specific heats,*** called the specific heat at constant pressure (the larger) and the specific heat at constant volume. The values of these are shown for a few gases in Table XIV, the values being, as before, the ratio of the number of calories required to heat one gram of the gas through 1° C. to the number required for a like change in a gram of water (which is one calorie). The ratio of the two specific heats of a gas is a number which is of interest in the kinetic theory of gases. Its value for a monatomic gas can easily be shown to be 5/3, and for a diatomic gas 7/5. Thus for air it is about 1.40.

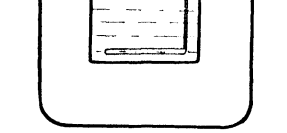


Fig. 12–4

A mixture calorimeter. The inner vessel is somewhat protected by the outer one.

The method of mixtures. The measurement of specific heats, and of other heat quantities, is generally known as ***calorimetry.*** A convenient experimental method of determining such quantities is furnished by the ***method of mixtures,*** especially if the highest accuracy is not required. This consists of "mixing" a hot body (or more than one) with a cold body (or more than one) and observing the resulting temperature. Figure 12–4 shows a suitable apparatus. Naturally, one of the bodies ought to be a liquid, or the mixture will not be likely to be a thorough one. A body of m grams mass, and of specific heat s, when raised through a ***change of temperature*** of $t°$ C. will absorb a number of

calories equal to mst. If the material is water the s may be omitted as it is very nearly unity. If, for example, a spoon of mass 50 grams, specific heat 0.1 and temperature 5° C. is put into a cupful (say 100 grams) of water at 90° C. (neglecting any action on the part of the cup or the air) the final temperature will be x degrees, where x is given by the "heat equation"

Heat given off by hot bodies = Heat received by cold bodies.

In this example mst for the spoon is $50 \times 0.1 \times (x - 5)$, and for the water $100 \times (90 - x)$; so that the equation becomes

$$50 \times 0.1 \times (x - 5) = 100 \times (90 - x),$$

whence it follows that

$$x = 86^\circ.0 \text{ C.}$$

If the water had been contained in a china cup (of 150 grams mass and specific heat 0.2), which would naturally always be at the same temperature as the water in it, the same equation gives

$$\underbrace{50 \times 0.1 \times (x - 5)}_{\text{cold spoon}} = \underbrace{100 \times (90 - x)}_{\text{hot water}} + \underbrace{150 \times 0.2 \times (90 - x)}_{\text{cup}},$$

and this yields $x = 86^\circ.9$ C. In the latter case we might say that the action of the cup is such as to add to the $100 \times (90 - x)$ calories lost by the hot water a quantity $30 \times (90 - x)$ calories lost by the cup; or that the cup is equivalent to 30 grams of water in its action. The product ms for the vessel (or any other object) is thus known as its ***water equivalent,*** or ***heat capacity,*** and is often useful.

It is obvious that experiments such as these could be done with the object of finding the specific heat of the cup, or of any other substance, if it were unknown; the final temperature would be observed and the unknown quantity could then be found from the heat equation.

The continuous-flow calorimeter. A somewhat more accurate method of calorimetry is known as the *continuous-flow method.* In one form of this experiment a stream of water flows through the "mixing chamber" or calorimeter in which the heat is being generated or received. It issues in a steady stream from the calorimeter at a temperature somewhat higher than that which it had when it entered. If the rate of flow of the water is measured, and the rise of temperature, the number of calories the water carries away per second is known. This must also be the rate of inflow of heat into the calo-

rimeter if it is not changing in temperature. Its temperature being constant, the calorimeter itself need not be considered in the calculation. This device is useful in finding the heat conducted down a bar (p. 176), the heat equivalent of work, the specific heat of gases, and in many other cases. In Fig. 12–5 a continuous-flow calorimeter is shown, arranged for finding the heat given by a Bunsen flame.

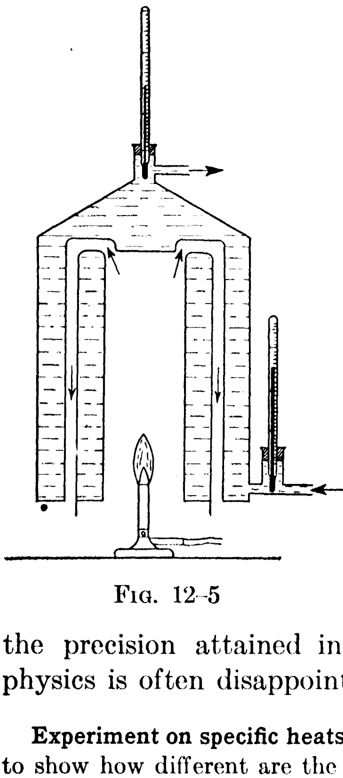


FIG. 12–5

The accuracy of heat experiments. In all calorimetric work it is very difficult to prevent heat from escaping to, or entering from the air, the hand of the observer, the thermometer, etc. If the experiment can be done inside a vacuum-walled flask so much the better, as heat transfer will thereby be cut down to a minimum. Otherwise, it is advisable to perform the experiment quickly, or to arrange matters so that the losses during part of the experiment are fairly accurately balanced by gains during the rest of it. In any case, however, the precision attained in this difficult branch of experimental physics is often disappointing.

Experiment on specific heats. An interesting experiment can be performed to show how different are the values of the specific heats of metals. If, for instance, specimens are made up of aluminum, iron, and lead, of equal weights, and these are heated to about 100° C. and quickly transferred to a large plate of paraffin resting on a ring stand, they will melt their way into the paraffin and through it if it is not too thick, *in the order of their specific heats.* To make the test a fair one each specimen should rest in contact with the same area of paraffin.

Examples. (1) Water is heated by means of paddle wheels driven by a 10 kgm. weight falling 2 m. If there are 160 grams of water in the vessel (the latter weighing 100 grams, and having a specific heat of 0.2) how much will the water rise in temperature?

The work done is 10×2 kgm.-meters = 20,000 gm.-meters. But 424 gm.-meters of work produce 1 calorie of heat. Hence 20,000/424 = 47.2 calories are generated. The water equivalent of the vessel is 20 gm.; adding this to the water we have 180 gm. in all, which will change by 47.2/180 or 0°.262 C.

(2) A hot coal is pulled out of a furnace and quickly plunged into a vessel (mass 200 gms., specific heat 0.1) containing 200 grams of water at 10° C. The

water rises to 35° C. The coal is later found to weigh 40 grams. If the specific heat of coal in this range of temperature be taken as 0.25, find the temperature of the furnace.

The heat equation is (calling x the desired temperature)

$$\begin{aligned} 40 \times 0.25 \times (x - 35) &= 200 \times (35 - 10) + 200 \times 0.1 \times (35 - 10) \\ &= 220(35 - 10) = 5500 \end{aligned}$$

whence $\qquad x - 35 = 550,$

or $\qquad x = 585°\text{ C.}$

PROBLEMS

1. A bathtub (mass 30 kgm., specific heat 0.1) contains 100 liters of water at 10° C. It is desired to warm it up to 35° C. How many liters of boiling water must be added, assuming no heat lost by conduction or otherwise? (1 liter = 1000 cm.^{3})

2. A 200-gram piece of iron is heated to 800° C. and dropped into 580 grams of water at 10° C. in a vessel whose water equivalent is 20 grams. The water is warmed 27° C. thereby. What is the specific heat of the iron?

3. A 100-gram china cup at 20° C. has 200 cm.^{3} of boiling water (at 100° C.) poured into it. The resulting temperature is 93° C. Find the water equivalent of the cup, and its specific heat.

4. A test tube (mass 20 grams, specific heat 0.2) contains 25 grams of water at 90° C. A thermometer at 10° C., whose water equivalent is 2 grams, is put into it in the attempt to read its true temperature. By how many degrees is the water cooled due to the presence of the thermometer?

5. How much heat would be needed to warm the air in a building whose volume is 5000 cu. m. (assumed to remain at constant pressure, 76 cm.) from 7° C. to 17° C.? (N.B. The air will expand, and some will escape only partly heated. From Table II (p. 16) and the laws of gases the average density can be found at the average temperature, and hence the mass heated can be obtained.)

6. 300 grams of brass wire at 100° C. are dropped into 285 grams of water at 10° C., in a vessel of 150 grams weight and specific heat 0.1. If the final temperature of the mixture is 17°.3 C., find the specific heat of the brass.

7. A house has brick walls of specific gravity 1.8; specific heat 0.2; thickness 25 cm.; total area 300 sq. m. If the inside of the house rises from 0° C. to 20° C., while the outside remains at 0° C., how much heat (in millions of calories) do the walls absorb?

8. A lead bullet weighing 20 grams is fired from a rifle with a speed of 200 m. a second against a hard wall. If half the energy of the bullet goes into heating it, how much will its temperature rise?

9. If a raindrop weighing 0.2 gram falls with a practically constant speed through a height of 1000 m., how much potential energy does it lose? If half this energy goes into warming the drop, how much will its temperature rise?

10. A 10-gram bullet is shot straight up into the air with an initial speed of 300 m. a second. It rises and falls again reaching the same point with a speed 100 m. a second, the loss being due to air friction. If one-tenth of this energy loss has gone into heating the bullet, how much warmer must it be on the return than at first? (The bullet is made of lead.)

11. A rock of specific heat 0.2 slides in an avalanche down a mountainside through a vertical height of 500 m. Assuming that half the energy lost is converted into heat in the rock, calculate its rise in temperature.

12. Ten blows are struck by a 700-gram hammer against a 100-gram mass of lead. Each time the hammer is moving at a speed of 200 cm./sec. just before it strikes. If all the energy goes into heating the lead, by how many degrees is it warmed?

Values of k for different substances are given in Table XV.

TABLE XV

Heat Conductivities. Solids and Liquids

Silver	1.096
Copper	1.000
Iron	0.167
Aluminum	0.50
Glass	0.002
Water	0.0014
Paraffin	0.0002
Hair felt	0.00009
Hard wood	0.0005

These quantities are often of considerable interest. If, for instance, one wishes to have a small space at a uniform temperature throughout, it is advantageous to make its walls of a good conductor, such as copper or aluminum. Such cases arise in cooking dishes, ovens, etc. In many cases low conductivity is desired, as for instance, in clothing, the walls of refrigerators or houses, the handles of frying pans, and the like. In a great many practical arrangements we need to calculate the amounts of heat transferred in this way, as, for instance, in the effort to conserve fuel in heating our houses in winter.

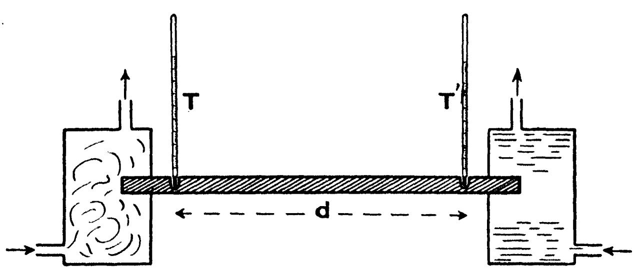


FIG. 13–1

Conductivity apparatus

Measurement of conductivity. A common laboratory experiment for finding the value of the coefficient of conductivity is arranged in the following manner. A thick bar of the substance to be measured (e.g. copper) is surrounded at each end by small chambers (Fig. 13–1). Through one of these steam is passed; through the other a continuous stream of cold water. Near the ends small thermometers are placed in holes in the bar, at T and T', whose distance apart is carefully measured. The conducting "plate" contemplated in the formula has then a thickness equal to TT', or d. The sides of the bar are wrapped with felt (not shown in the figure) in an effort to prevent transverse losses of heat. The rate of flow of water through the water chamber and

its rise of temperature between entering and leaving are measured. From these H is obtained, and the dimensions of the apparatus give us A and d. The number of seconds n is the same as the time over which the water flow is measured. Hence all the quantities in the formula are known, and the coefficient may be found.

Heat conduction in the earth. A problem involving heat conduction, which is of great interest and on a grand scale, is furnished by the earth itself. It is found by descending into mines that the temperature rises by about 1° C. for every 30 meters depth (1° F. for 60 feet). This state of affairs demands a steady flow of heat from the interior of the earth to its surface. Given the conductivity of the materials of the earth's crust, and the variations of temperature with depth, it is an interesting mathematical problem to find the rate at which heat is being lost from the whole earth, and to try to calculate the rate of cooling of the earth and hence estimate the time back to the age when the earth's crust was molten. Unfortunately there are too many unknown quantities involved, and the age of the crust which was thus derived (roughly 100 million years) has been discarded as of little interest since radioactive materials (Chap. 39) were found to be supplying an unknown and perhaps large amount of heat on account of their continual disintegration.

Convection of heat. Liquids and gases when heated expand and become lighter than before. Unless the heating is uniform throughout, hydrostatic forces start convection currents in the material, whereby the cooler parts sink and the warmer are forced up. This produces a circulation of heated material, which is an example of the commonest and most effective mode of heat transfer. The household "steam radiator," for instance, (better called a "convector") radiates some heat, as we shall see, but loses most of its heat to the room by convection of the air heated by contact with the pipes (conduction) and streaming continually upward past them. The draught in a chimney is of this type, and will carry away a large part of the heat of the fire unless one makes plans to prevent it. A stove in a room with a long, exposed stove pipe may not be an object of great beauty, but it heats the room much more effectively than a fireplace, in which all the convected heat is lost up the chimney. Convection is the main cause of winds. The sun's heat warms the lower layers of the atmosphere; these are then forced to rise, and a circulation results. At the

tropics this occurs steadily on a grand scale, so that the cooler air flowing in to displace the lighter heated air gives rise to the constant "trade winds" as shown in Fig. 13–2. Convection is of use in the warming of houses by hot-water or hot-air systems, and in the cooling of automobile engines by the circulation of water. The smoke rising from a cigarette usually discloses a very complex example of convection in the air, and on a larger scale the same sort of irregularities in air temperatures and air currents are the cause of the "holes" or "bumps" in the air, which are the bane of aviators.

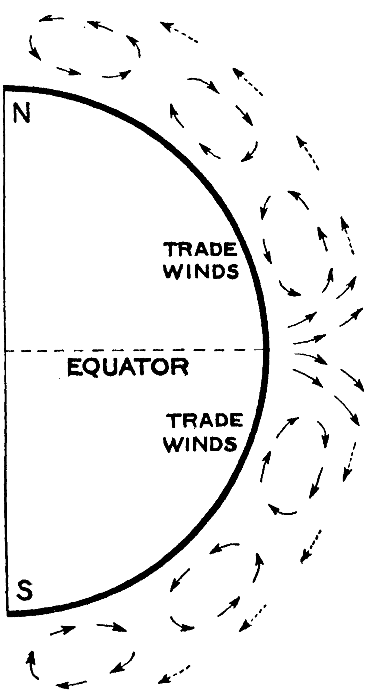


Fig. 13–2

Simplified scheme of average atmospheric circulation. Several features of this plan, including the upper current direct from equator to pole, are not yet definitely established.

Radiation. The third mode of heat transfer is a much more mysterious matter. If one brings a hand near a kettle full of very hot water, a stream of warm air is found to be rising from it, but beneath it, or at the side, heat may also be felt, which is due to a different process. A sheet of cardboard, or better of metal, will intercept this heat; it is invisible; it appears to travel instantly, and in straight lines. If a hot body is placed in an exhausted glass vessel, like the luminous wire in a small incandescent electric lamp, the heat comes out without apparent difficulty through the (nearly) empty space. The same process brings us heat from the sun, or even from the much more distant stars, across vast spaces which we know are completely devoid of all but wisps of ordinary matter. The human mind finds it hard to conceive of any mechanical action taking place across a space that is absolutely empty. Two alternatives suggest themselves; either radiation is like a stream of small bullets, or corpuscles, shot out from the radiating body, or there is in all space some sort of a non-material medium (whatever that means) which we refer to

as ***the ether,*** capable of carrying energy from place to place in the form of waves. The corpuscular theory was supported by Newton, was later discredited, and is now being partially revived, in combination with wave ideas, in the quantum theory (p. 545). The wave theory at the present time seems to fit in with more facts than any other, and we shall use it at considerable length in later chapters (31 and 35). If the radiation of heat is a wave process, we must imagine a medium in which these waves occur. Filmy bodies such as comets appear to pass through this medium without friction; the planets revolve about the sun without being retarded in their regular paths; hence the ether must be a frictionless medium quite unlike any known material. Not only does the radiation of heat occur in it, but electric and magnetic effects also, and light. While it seems at first sight as though the ether must have very complex properties to be able to do so much, in reality only one kind of action is required of it. All the phenomena just mentioned are closely related. This does not, of course, make that action comprehensible, and in the present state of physics we cannot "explain" it. Because of this, or for other reasons, the ether is regarded by some scientists as non-existent. Perhaps they are right, but it is so great a convenience in helping us to correlate a variety of phenomena that it will be treated in this book as real, at least whenever it is useful.

The ether waves which produce, or constitute, radiation of heat have different lengths and rates of vibration. These will be considered later on, after the nature of waves has been examined. (See especially pp. 534, 555.)

Effect of surfaces on radiation. The amount of energy escaping by radiation from a surface depends on the nature of that surface. A cubical box, filled with very hot water, is often used to show this effect; one side is highly polished, another blackened. A thermometer, preferably with blackened bulb,[1] can be put near the box, and will show that it is receiving more heat when the black side is near it than when the shiny side is the same distance away. In general, rough surfaces radiate more than smooth ones at the same temperature, and absorb incoming radiation better. It is interesting to test the influence of the surface of a thermometer on its readings when it is being heated by radiation. Two thermometers can be used, one blackened, the other untreated. They

[1] Use lampblack in alcohol with a little shellac.

will record very different temperatures if placed together in sunlight.

Black bodies. Stefan's law. An ordinary blackened body may be very black to light (that is, it may completely absorb it), but not to heat radiations.[1] An ***absolutely black body*** is one that absorbs all the radiation that falls on it; these bodies also radiate more than any others at the same temperature. Such a body can be made in the form of a hollow chamber with blackened walls with a small hole in one side. Any radiation entering the hole will be absorbed, and such a body, if uniformly heated all over, will radiate more per unit area (through the hole) than any other body at the same temperature; or, if heated to a red heat, it will be brighter than any other body at the same temperature. An Austrian physicist, J. Stefan, formulated a law in 1879, now amply

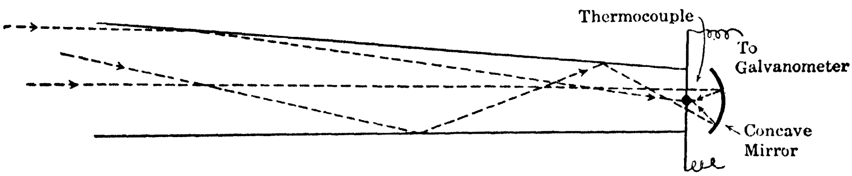


Fig. 13-3

A radiation pyrometer. A conical reflector receives radiation from a furnace and concentrates it upon a receiving thermopile at the small end. The electric current from the thermopile gives a direct temperature reading on a calibrated voltmeter.

verified over a wide range, connecting the heat, H, radiated (per second per square centimeter) by a black body with the absolute temperature. It is $H = cT^4$, c being a constant whose value is 5.72×10^{-5}, if H is in ergs, and T, as usual, is the absolute temperature. Ordinary bodies radiate less, but follow the same law, and the "constant" c has in such cases a smaller value, and varies somewhat with the temperature. If one has any accurate way of measuring the heat received from hot bodies (furnaces, for instance, in an industrial plant) one may use Stefan's law, and find the temperature of these bodies by the radiation received from them.[2] This is the basis of one form of "radiation pyrometer,"

[1] All sorts of radiation, such as light, X-rays, etc. carry energy with them, and produce heat where they are absorbed. "Heat radiation" does this more than the other sorts and can do comparatively little else; hence the appropriateness of the name.

[2] Strictly speaking, the heat received is $H = c(T^4 - T_0^4)$, where T_0 is the temperature of the measuring instrument, which is itself radiating at the same time.

which is actually used[1] (Fig. 13–3). It is commonly applied to furnaces, which are hollow bodies with rough walls, observed through a small hole, and are therefore examples of nearly "absolutely black" bodies, to which Stefan's law accurately applies. As the law was also deduced theoretically by Boltzmann of Vienna, we have confidence in its validity, and should expect it to hold for temperatures beyond the range in which they can be tested against the gas thermometer. Thus it can be applied to finding the temperatures of bodies such as the positive (carbon) pole of a direct-current arc (3500° C.), or even the sun itself (5700° C.).

Exchange of heat by radiation. Stefan's law implies that all bodies, even cold ones, radiate heat if their temperature is above the absolute zero. A piece of ice in a warm room radiates heat, but the other objects in the room radiate more heat. Hence the ice exchanges heat with the other bodies, receiving more than it gives out. With a sensitive instrument (such as a thermopile with thin blackened junctions, associated with a sensitive galvanometer; see p. 335) one may measure the heat received by radiation from a man at a considerable distance. If a piece of ice is substituted for the man, the opposite effect is noted, and an observer might be tempted to call this the radiation of cold. Cold is, however, merely the absence (or loss) of heat. In such an experiment, the idea of heat exchange shows us that the measuring instrument radiates heat in both cases, but receives more from the man, and less from the ice than it sends out. Hence it is warmed by a warm body, and cooled by a cool one, in both cases by the radiation of heat.

Solar radiation and atmospheric absorption. It is of vital importance to human life that we receive a steady supply of heat from the sun. This all comes to us in the form of radiation, some of it visible as light, some of it in invisible rays. The sun sends to each square centimeter of surface placed at right angles to its rays an amount of heat equal to 1.938 calories per minute. This is not, however, all received at the surface of the earth; a variable portion, often as much as one calorie, is taken out of it by absorption, scattering, etc. in the earth's atmosphere. Even on a clear day the absorption may be considerable. At times when a violent volcanic eruption occurs somewhere on the earth, fine dust is thrown so high into the air, and in such quantities, that it may take a year or more to fall, and during this time it fills the atmosphere of the entire earth. This happened in 1783 when Asama (Japan) exploded; again in 1815, when Mt. Tambora on Sumbawa Island (East Indies) is said to have

[1] Invented by Dr. C. B. Thwing, to whom we owe this diagram.

killed some 56,000 people, and produced darkness for three days over a distance of 300 miles; and again in June, 1912, when the eruption of Katmai (Alaska) occurred. The fine dust driven into the upper air scatters some colors more than others (p. 572) and produces more brilliant sunsets than usual. Its most important effect, however, is a reflection back into outer space of a larger part of the sun's heat than usual, which causes a fall of temperature over the whole earth's surface after any considerable volcanic activity. The year 1816 is famous as "the year without a summer"; the years 1783–5, 1884–6, 1912, and others, show the same effect, directly traceable to violent eruptions.

It has been suggested on account of these facts, that the great glacial periods, when the climate of the whole earth seems to have become much colder than normal, might have been due to long-continued volcanic disturbances.[1] It might be noted in this connection that the sun's heat is itself slightly variable. From some cause or other, possibly tides in the sun created by the attraction of the larger planets, the sun's surface becomes periodically disturbed

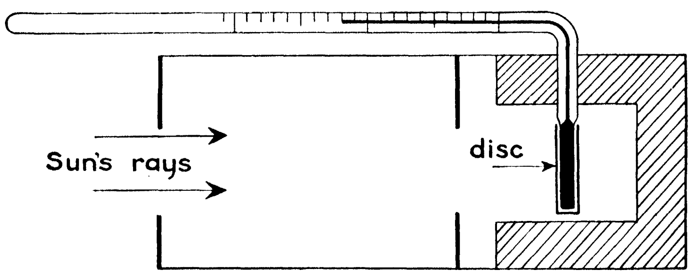


FIG. 13–4
Abbot's pyrheliometer

and huge storms occur, which appear to us as sun-spots, varying in their frequency with an approximately eleven-year cycle. The heat sent to us by the sun also varies slightly for the same reason, and with the same period, being greater at the times when the sun-spots are most numerous. The change is small, and appears to have no marked effect upon our climate.

Measurement of the heat received from the sun. The heat received at the earth's surface from the sun is measured by instruments called "pyrheliometers." In Ångström's form of this instrument, the sun's rays are allowed to fall on one of two thin, similar, blackened strips of metal (the alloy manganin is used). The other is shielded from the radiation, but warmed by an electric current to the same temperature. Two very small thermocouples (p. 367) are used, almost in contact with the backs of

[1] A full account of these matters, along with much else that the reader will find interesting, is to be found in "The Physics of the Air" by W. J. Humphreys, 2nd edition, 1929 (McGraw-Hill).

the metal strips, to tell when these differ in temperature, and this may be done with great precision. Electrical measurements (p. 372) enable us to measure the heat delivered by the current, and this must then be equal to that received from the sun.

In Abbot's[1] form of this instrument, the sun's rays are allowed to fall on a blackened silver disc, thick enough to contain inside it a very small mercury thermometer bulb, as shown in Fig. 13–4. The disc is in a chamber which has heat-insulating walls, and this

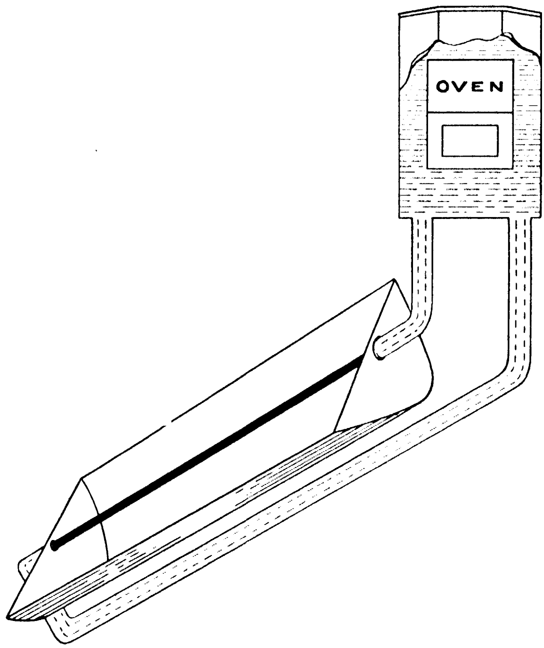


FIG. 13–5

Dr. Abbot's solar cooker. A large parabolic reflector concentrates the sun's rays on a blackened tube containing oil, which heats the oven by convection.

is so mounted that it can be made always to face the sun. Diaphragms limit the area of the beam admitted, so that it just covers the disc. The thermometer stem is bent at right angles for convenience. After the instrument is calibrated, the heat received from the sun can be found from the time taken for a certain rise in temperature.

Power from the sun. In sunny climates, the heat of the sun has been gathered up by large reflectors, and used for cooking purposes (Fig. 13–5), or even to heat water for a steam engine. There

[1] Dr. C. G. Abbot, Secretary of the Smithsonian Institution at Washington.

is available from this source a great deal of power, but the necessary apparatus is inconvenient and expensive, and in most climates sunshine is too intermittent to be dependable.

Books recommended:

Abbot, C. G., "The Sun," 1911 (Appleton).

Ingersoll, L. R., and Zobel, O. J., "Heat Conduction with Engineering and Geological Applications," 1913 (Ginn and Co.).

Steinmetz, C. P., "Radiation, Light, and Illumination," 1910 (McGraw-Hill).

PROBLEMS

1. A room is at a temperature of 20° C. when the air outdoors is at − 10° C. Is it reasonable to say that these are also the temperatures of the inner and outer surfaces of its glass windows? Explain.

2. A vessel is arranged in two parts, separated by a wall of aluminum whose area is 100 cm.^{2} and thickness 1 cm. One side is kept at 100° C. by a steam supply, and the other at 40° C. by means of running water. If 144 grams of water flow through this chamber per second, and come out 20° C. warmer than they entered, find the conductivity of the aluminum.

3. A kettle made of copper 0.5 mm. thick has a total outside area of 2000 cm.^{2} If it is filled with water at 100° C. and its outer surface is at 90° C., how many calories will flow through its wall per minute?

4. A house has brick walls of conductivity 0.01, thickness 25 cm., and total area 200 square meters. If the difference of temperature of the inside and outside of the walls is maintained at 10° C. in winter, how much heat (in millions of calories) do the walls conduct outward per hour?

5. If plate glass (thickness 7 mm.) is used in the windows of a room (total area of glass 5 sq. m.) instead of common window glass (thickness 3 mm.) and if the temperature of the inside of the glass is 10° C. higher than that of the outside, how much saving of heat per hour will result from this change? (Assume $k = 0.0025$.)

6. If a furnace gives a certain amount of heat radiation to an instrument capable of measuring it, when the furnace is at 827° C., what will be its temperature when it gives twice, four times, and sixteen times as much? (The temperature of the receiving instrument may be neglected in comparison with that of the furnace.)

7. A kettle full of water at 97° C. radiates heat to my hand which has a surface temperature of 27° C. How much more would a block of ice of the same size receive from the kettle when placed in the same position? Why?

8. If black bodies are those that absorb more heat at a given temperature than any other sorts of bodies, show that they must also emit more. (Assume two bodies, one black, the other not, in an enclosure at a uniform temperature.)

CHAPTER 14

CHANGE OF STATE

Latent heat, 185; measurement of latent heat, 186; measurement of latent heat by cooling curve, 186; undercooling, 186; latent heat effects, 187; melting point, 188; solutions, 188; freezing mixtures, 189; the boiling point, effect of impurities, 189; osmosis, 190; boiling a cooling process, 191; boiling and freezing at low pressures, 191; refrigerating machines, 192; saturated vapor and its pressure, 192; gases dissolved in liquids, 195; humidity, 196; hygrometers, 196; dew, fog, and clouds, 197; graphical representation of properties of water, 198; pressure and melting point, 198; ice formation in rivers, 200; isothermal curves, the critical point, 200; "dry ice," 203; the liquefaction of gases, 203; apparatus for liquefying air, 204; the liquefaction of other gases, 205; experiments with liquid air, 205.

Latent heat. Interesting heat effects are associated with melting, freezing, boiling, evaporation, and condensation, all of which are classified as phenomena involving a change of state.

If a glass full of ice and water is watched, the proportion of ice to water is usually found to diminish because the warm air of the room is continually supplying heat to the mixture, but the temperature of the mixture does not change if it is kept well stirred. Even if heat is supplied to such a mixture at a much more rapid rate, by a hot stove, for instance, the temperature remains constant so long as there is ice enough to make thorough stirring effective. Again, if the temperature of a kettle of water on a stove is watched, it is seen to rise to a fixed value, and then to remain constant while the water is boiling away. In all such cases heat is entering the substance continually, and yet the temperature is not changing. This heat which does not show itself is said to be ***latent heat,*** and is, of course, connected with the change of state which is going on. It takes work to overcome the forces holding the particles of the ice, or of the water, together. This work is furnished by the heat which is being supplied, and this is stored as potential energy in the resulting water or steam. The very large magnitude of the molecular forces to be overcome explains the large value of the latent heat which we find when we come to measure it.

Measurement of latent heat. The numerical value of the latent heat associated with a change of state of a substance is taken to be *the number of calories required to change one gram of it without changing its temperature.* The measurement of this quantity is usually carried out in the laboratory by the method of mixtures. The latent heat of melting ice, for instance, is found by putting a weighed amount of ice at 0° C. into warm water. The resulting fall of temperature in the water is measured. On one side, the warm water and its container lose heat and fall in temperature. On the other side the ice gains mL calories in changing from ice at 0° C. to water at the same temperature, (m being the mass of ice, and L its latent heat) and then the water resulting from the melting takes still more heat in rising to the final temperature of the mixture. Hence a heat equation can be formed and solved for the latent heat. This process is illustrated by examples below.

A somewhat similar experiment in which steam is delivered into cold water yields the latent heat of steam, which is the number of calories required to turn one gram of water at 100° C. into steam at the same temperature, and at standard pressure (76 cm. of mercury).

Measurement of latent heat by cooling curve. By the simple expedient of watching the temperature of a molten substance while it solidifies, and timing it, the latent heat may be found. For instance, if m grams of molten lead be allowed to cool, and the rate of cooling observed, the lead is found to lose a quantity we may call mst (p. 171) calories per second, where s is the specific heat of the liquid, and t the fall of temperature per second. If the temperature and time are then followed, and if the lead is pure, the temperature will be found to remain steady at the melting point, while the metal solidifies. All this time it is just as high in temperature above its surroundings as it was, and is therefore losing heat at the same rate, so long as the surroundings also remain at constant temperature. If the rate of heat loss is known, and N is the number of seconds during which it remains at constant temperature, a quantity of heat equal to $mst \times N$ is lost in this time, and this must be the heat given out in solidifying, and be equal to mL. Hence, the latent heat L can be found. If the temperature is plotted against the time, a "cooling curve" is obtained, as in Fig. 14–1, and this shows a level portion AB, whose ordinate OT gives the temperature of the melting point.

Undercooling. In some cases, and especially with certain substances (e.g. naphthaline, or acetamide), the cooling curve may take the form shown in Fig. 14–2, where the portion AKF represents the phenomenon of undercooling. The liquid follows the usual cooling curve DA down past the temperature T, which is

its proper melting (and freezing) point without the appearance of any solid. There seems to be some difficulty connected with the formation of the first bit of crystal. While it is in this undercooled condition, jarring is sometimes sufficient to start solidification, and it can always be started by dropping a small crystal of the same substance into the liquid. When solidification begins, a large quantity changes its state at once, and the temperature immediately rises from the level of K to F on the line AB, at the true freezing temperature. This rise seems very odd, but is due to the liberation of latent heat, from the form of potential energy in the liquid to kinetic energy in the solid. It shows that freezing is really a warming process. The experiment of undercooling

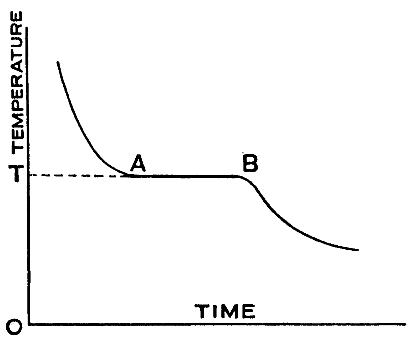


FIG. 14–1
A cooling curve for a pure substance

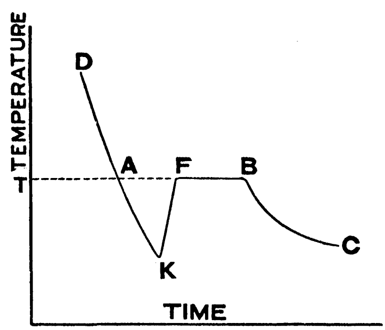


FIG. 14–2
A cooling curve showing undercooling

has been carried out with water to the extent of 20° below 0° C., though 5° is about all that can usually be managed.

Occasionally in winter water drops may fall as rain through layers of air cold enough to undercool them, whereupon they freeze instantly when they touch anything. This explains the fact, sometimes noticed after an "ice-storm," that the ice coating formed on every twig is mainly on its upper side.

Latent heat effects. Table XVI shows that the latent heat of water is a large quantity. The effect of this upon climate is noteworthy. In the presence of large bodies of water, sudden changes in the temperature of the air are retarded on account of the large specific heat of water; but if the air temperature crosses 0° C., the retarding effect is greatly increased by the liberation of latent heat by freezing (if the temperature is falling) or by the absorption of it by melting (if the temperature is rising). A barrel of water

TABLE XVI

Latent Heats of Melting and Vaporization

(in calories per gram)

	Melting	Vaporization
Water (ice, steam)	80	540
Mercury	3	68
Tin	14.6	—
Lead	5	—
Ammonia	108	341
Sulphur dioxide	—	96
Liquid air	—	50

in a fruit cellar acts likewise as a reservoir of heat to keep the temperature of the room from falling far below the freezing point.

The latent heat of steam is much larger than that of ice; part of this (40 calories) is the work that must be done to push aside the atmosphere, and make room for the vapor.

The latent heat of evaporation of water at a temperature lower than 100° C. is larger. Thus it requires nearly 600 calories per gram to turn water at 0° C. into vapor at the same temperature.

Melting point. The freezing or melting point of a substance is defined as the temperature at which its solid and liquid forms can exist together in equilibrium, without any change in their relative amounts. At this temperature neither freezing nor melting can occur, unless heat is taken away from or added to the material. This temperature is a very definite one for pure substances, especially those that solidify in a crystalline form. Mixtures, like ordinary wax, glass, or tar, pass through a plastic stage with no definite melting point. The flatness of the portion AB in the cooling curve (Fig. 14–1) is often an accurate test of the purity of the substance, since impure ones give a rounded, sloping line.

Solutions. Solutions exhibit some curious thermal effects. To begin with, some substances when dissolved in water absorb heat and lower the temperature of the solution, while others act oppositely. Liquids and gases (e.g. ammonia) when dissolved in water evolve heat; solids have a "heat of solution" which may be either positive or negative; caustic potash, for instance, evolves heat, while common salt and other salts absorb it.

If a non-volatile substance is dissolved in water and the solution is cooled, it is found to have a lower freezing point than pure water has, and the freezing occurs in a somewhat different manner.

When crystals of any sort are formed, they are made up of particles which fit with one another into a certain characteristic shape; foreign particles are rejected, just as a mason would reject a round brick in building a wall. When salt water begins to freeze, crystals of pure ice are formed, and the salt is left in the solution, which thus becomes more and more concentrated, until the liquid can hold no more salt. At this temperature (− 22° C. with common salt) the solution is said to be ***eutectic,*** and if it is cooled beyond this point, salt crystals and ice are formed side by side in a fine-grained mixture.

These facts can be more easily remembered if we make the convenient (but perhaps not entirely satisfactory) assumption that the salt and the water particles attract each other, and tend to stay together as a solution even at temperatures so low that, in the absence of salt, the water particles would gather themselves up into solid form.

Freezing mixtures. If dry salt crystals and cracked ice, both at 0° C., are mixed together, the assumed attraction between the salt and water particles is sufficient to pull the particles away from one another in the solid form, so that they may come together in the solution. But, to break down the solids, work must be done and energy used up. Since no external energy is supplied, it is taken from the kinetic energy of all the particles; their agitation is decreased and their temperature falls. Hence such mixtures are called ***freezing mixtures.*** With suitable materials (for instance, crystallized calcium chloride) temperatures as low as − 50° C. are reached in this way.

The boiling point. Effect of impurities. The ***boiling point*** of a liquid is defined as the temperature at which the liquid and its vapor can exist together at atmospheric pressure in equilibrium, without one changing into the other unless heat is taken in or given out. It can best be measured by placing the thermometer in the vapor, above the boiling liquid. The pressure in the vessel must be noted (see p. 194).

Salt added to water produces a change in the temperature of the liquid when boiling, which is explained in terms of the attraction already postulated between salt and water particles. This force makes it more difficult than usual for the water particles to escape from the liquid into the vapor; hence the temperature of the liquid is raised by the presence of dissolved salts. This rise may amount

to several degrees, as shown in Fig. 14–3, and is often of practical importance. For instance, in cooking, salt added to water makes possible a higher temperature, and shortens the time required. The vapor above a salt solution is, however, composed of pure water, and will always be at 100° C. if the pressure is normal (76 cm.). A thermometer placed in this vapor, above water containing dissolved salts, will show the true steam temperature unless spattering covers it with some of the solution, in which case it will indicate too high a reading. If the impurity in the water is not salt but some substance easily vaporized, the effects are different; with alcohol, for instance, the vapor is a mixture and the temperature of the boiling liquid is lowered. The kinetic theory explains this as due to the action of the particles in the vapor (p. 194).

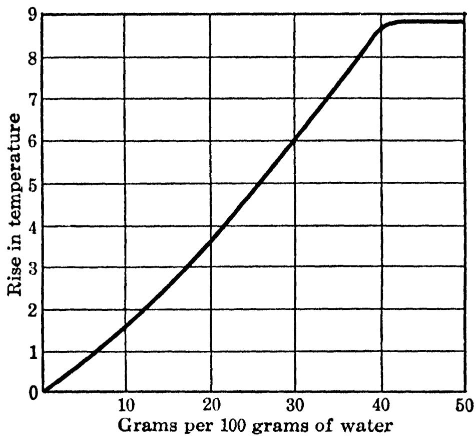


Fig. 14–3

The rise in the temperature of boiling water due to added salt

Osmosis. Another phenomenon which is closely related to those just considered is that of ***osmosis.*** It consists of the diffusion of water, certain solutions, or gases, through membranes which permit simple molecules to pass, but do not admit more complicated ones. For instance, a dried raisin soaked in water swells out to its original size, or even more. Its skin allows water to enter, which it does perhaps because it is attracted by the sugar inside, but the sugar itself cannot easily escape. A membrane like this is said to be ***semi-permeable.*** If a "thistle tube" (an open tube enlarged at one end) is filled with syrup and its large end, covered with a suitable membrane (such as pig's bladder), is held under water, the water will enter in order to mix with the sugar, and the level inside may rise to a great height. The resulting excess of pressure on one side of the membrane is called ***osmotic pressure.*** It is a curious fact that this pressure varies according to the laws already familiar to us as the gas laws ($pv = RT$, p. 160); the volume in these relations now being understood to be the volume of water

throughout which a given mass of the dissolved substance is diffused. Osmotic pressures may be very large, often as much as a thousand atmospheres.

Osmosis is a very important process in our own bodily systems. For instance, dissolved gases pass continually in and out of the blood in our lungs through membranes; also, the products of digestion diffuse by osmosis through the walls of the alimentary canal.

Boiling a cooling process. According to the kinetic view of the nature of the boiling process, it should occur much more easily if the atmospheric pressure is partly removed. This is easily shown. A glass of luke-warm water will boil vigorously if placed under a bell jar connected to a good vacuum pump, and if its temperature is watched, it will be found to be falling while the boiling lasts. An ordinary kettle, boiling on a stove, is keeping itself from getting any hotter by sending off the incoming heat in a latent form in the vapor which is generated. The boiling water is thus enabled to remain at a constant temperature. Hence boiling may be regarded as a *cooling process.* In terms of our kinetic theory, the particles most likely to escape from the water surface are those that happen to have at the moment the greatest speed. Hence they carry off with them an undue proportion of energy, leaving the liquid with less (i.e., at a lower temperature) unless the supply is renewed. In the case of the glass of lukewarm water boiling in a partial vacuum, no heat is being supplied to keep the temperature constant, and therefore the temperature must fall while the water boils.

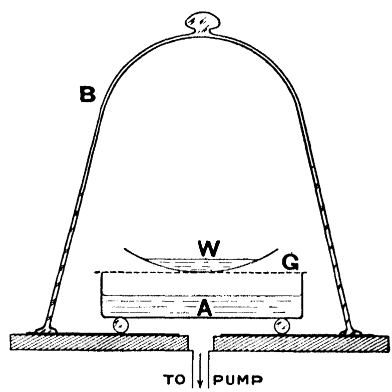


Fig. 14–4
Freezing water by boiling it in a vacuum

Boiling and freezing at low pressures. If the water is ice-cold, it may still be made to boil, provided the pressure is reduced to a very low value; and in boiling it will again lose heat. But, if it is already at 0° C. and loses heat, part of it must freeze. Hence we may have water boiling and freezing at the same time. The temperature and pressure required for this are: $p = 0.46$ cm.; $t = 0°.0075$ C. This condition is known as the "triple point" (p. 198). The experiment requires some sort of pumping device which will dispose of water

vapor. With the usual types of vacuum pump, the air under the bell jar, B, may readily be removed (Fig. 14–4), but water vapor is likely to be expanded and evaporated in one part of the pump stroke, and compressed and condensed in another, so that it remains, and the pump may become loaded with water so that the pressure will go down no farther. This experiment succeeds best by putting a small quantity of water, W, ice-cold to begin with in order to save time, in a flat dish supported on a gauze, G, over a vessel partly filled with concentrated sulphuric acid, A, which effectually absorbs water vapor, and keeps the pressure down. A thick sheet, or ring, of soft rubber between the bell jar and the base-plate will ensure an air-tight contact, without which the experiment will not succeed.

Refrigerating machines. Much interest is now shown in the manufacture of "artificial" ice, and in domestic refrigeration. Most of the devices used for this purpose work on the following plan. A "working substance" is used which can be liquefied by pressure at ordinary temperatures. This is done by compressing it in a tank, or coil, by means of a pump run by an electric motor, and carrying off the heat produced in this process by either air or water cooling. The resulting liquid under pressure is allowed to expand and boil off in a closed space where the necessary latent heat is ***not*** supplied, and is therefore "stolen" from the surroundings. These are thus cooled to a low temperature. It is often arranged that the substance cooled is a concentrated brine which does not freeze, and this circulates through a piping system passing through the refrigeration space. The success of such devices depends on theoretical as well as practical considerations. Theoretically, the substance with the largest latent heat is the best, but practically, it should be one that does not require excessive pressure to liquefy it, nor should it be corrosive or poisonous. One of the substances that meet these requirements best, among those that are easily available, is sulphur dioxide; this requires only a moderate pressure, and appears at first to be quite satisfactory, until one finds that it is a poisonous gas, and that it combines with water to form a corrosive acid. In spite of these difficulties, very successful machines are in common use, with sulphur dioxide as the working substance. One is shown in Fig. 14–5. In larger plants ammonia is often chosen.

Saturated vapor and its pressure. Ordinary evaporation consists of the escape of the faster-moving particles from the surface of a liquid, accompanied by cooling, as explained above. In a closed, exhausted vessel, the liquid will evaporate to a certain extent,

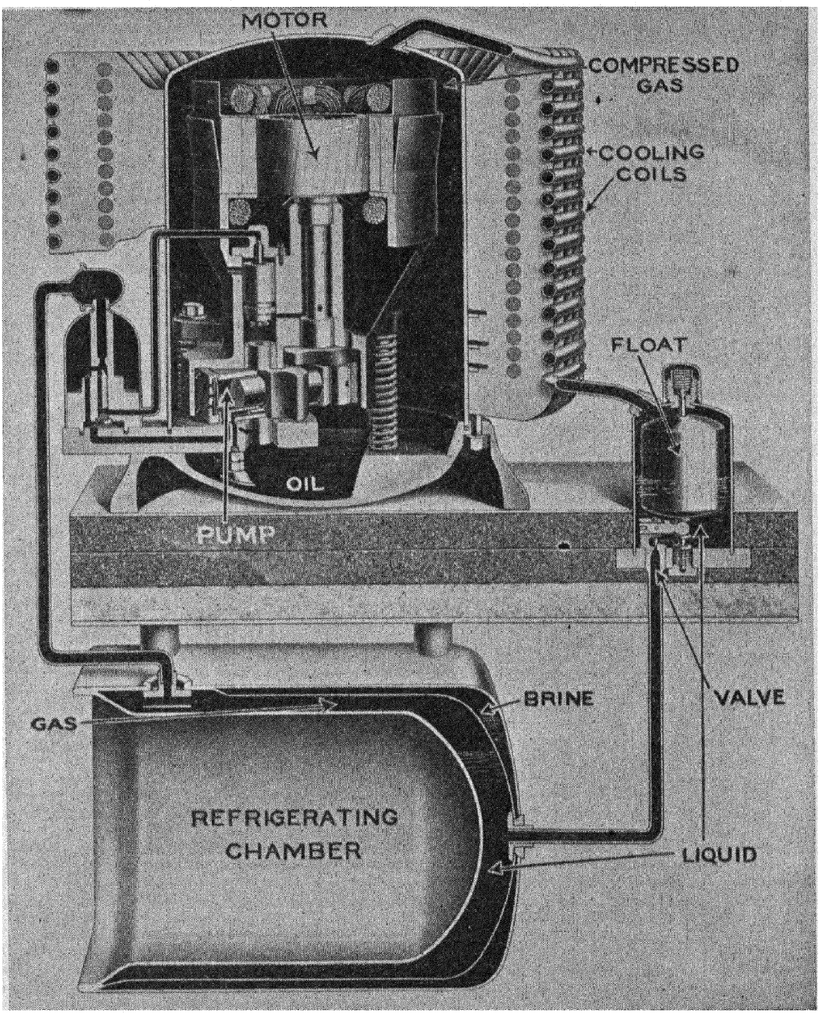


FIG. 14-5

A domestic refrigerator. The liquid evaporates inside the shell surrounding the refrigerating chamber, and is then compressed by the pump into the space in which the motor runs. From here the compressed gas goes through the cooling coils and drops as liquid into the float chamber where its return flow is regulated. (Courtesy of the General Electric Company.)

"filling" the space above with vapor particles, which act like the gas particles of the kinetic theory. Some of these will strike the liquid surface again, and be recaptured, and this will be the more likely to happen the more vapor particles there are. Hence there will be a limit, in such a closed space, to the number of vapor particles which can exist at any moment in the free state. When this limit is reached, the space is said to be "saturated" with the

vapor, and the pressure in it is called the ***saturated vapor pressure.*** If any more were added to it, the extra amount would promptly re-condense. A rise in temperature enables more to escape, without increasing the chance of recapture at the same rate; so that the pressure of the saturated vapor increases with the temperature. A very good type of ***thermometer*** has actually been made which is based on this idea. At the "boiling point" of any liquid the pressure of the vapor has risen so that it is equal to that of the atmos-

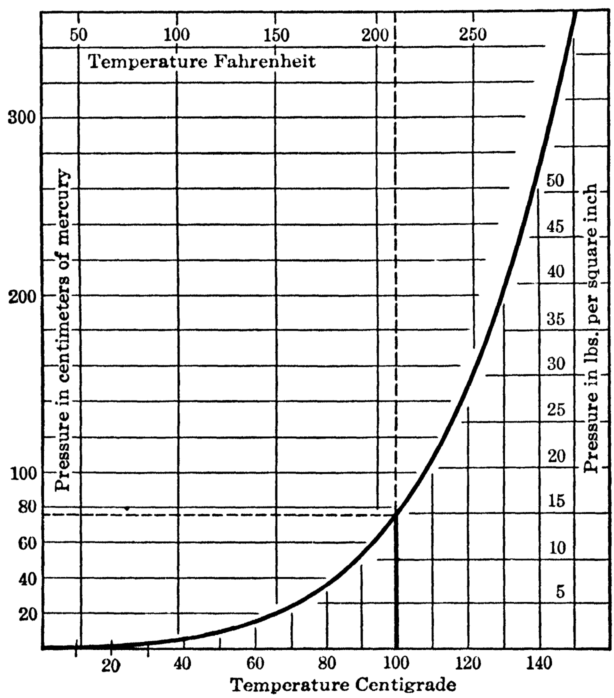


FIG. 14-6
Curve of saturated vapor pressure for water

phere, and the vessel could then as well stand open to the air. In fact, this condition may be taken as defining the boiling point. The boiling point of water is affected by the addition of a volatile impurity like alcohol because the alcohol particles escape readily and help the water particles to produce a vapor pressure equal to that of the atmosphere at a lower temperature than would be possible for water alone.

If the saturated vapor pressure is plotted against the temperature, a curve such as that shown in Fig. 14-6, is obtained. It is

to be noted that this pressure is that contributed by the vapor only. If air is present in the vessel at the same time, the liquid will evaporate into the space "occupied" by the air to nearly the same extent as though the air were not there, and the total pressure inside will be the sum of the two pressures, that of the air, and that of the saturated vapor. The chief difference caused by the presence of the air is that the air particles get in the way, and prevent rapid evaporation, so that it may take minutes, or even hours for the space to become saturated, instead of seconds as it would if no air were present. The air particles, when they strike the liquid surface, do not adhere to it to any considerable extent; they therefore remain in the space above, and continue to exert an undiminished pressure. When such a space contains all the water vapor it can hold, we commonly say that "the air is saturated." Strictly speaking, it is the ***space*** that is saturated; the air plays an inactive rôle.

Gases dissolved in liquids. A small proportion of a gas in contact with a liquid may, it is true, dissolve in it, and permeate it, as happens to a large extent with carbon dioxide gas (in "soda water"), or ammonia, especially under great pressure; but this effect is small in the case of water and air. It is easily observed, however, when large pressures are involved. If water is drawn out of a faucet from a high-pressure system, air sometimes appears in it as a cloud of fine bubbles, giving it a milky appearance. This solution of air in water is also disclosed if one raises the temperature of ordinary water toward boiling. The first bubbles that quietly grow on the walls of the vessel, before it has begun to "sing," are air bubbles coming out of solution; the air is not so well retained at high temperatures, and boiling removes practically all of it. Water also gives up its dissolved air on freezing, making the familiar bubbles that we see in ice.

Air is also found to cling to the surfaces of many solids in the form of a thin film which may make itself troublesome. It often happens in physical experiments that a tube containing metal and glass parts is to be evacuated. If an extremely low pressure is desired the whole tube must be baked at a temperature as high as possible in order to "outgas" its parts. Apparently a metal may hold a considerable amount of such a gas as hydrogen in its interior, which cannot be removed by any less violent treatment. Glass retains water in like manner.

Humidity. The air we breathe is a mixture of many constituents. One of these is water vapor, and usually far less of this is present than would be if the air were saturated. The ratio of the amount present to that which would saturate the air is called the ***relative humidity,*** or often simply the humidity. As the amount that will saturate the air increases rapidly with the temperature (see Fig. 14–6) there is an important difference between the amount of moisture in the air and its humidity. To illustrate: on a damp winter's day, the air may be saturated with water vapor. It may be drawn into a house and heated, without any water being added to it, or taken away. It then becomes warm air which is capable of taking in far more water vapor than it has; that is, it is now what we call very dry air. Thus damp air has become dry air without changing the amount of water it contains. This shows why the air of so many American houses is dry in winter; their humidity is often less than that of the air of a hot desert. This low humidity dries and cracks furniture and causes excessive evaporation in plants, etc. It probably does harm to our nasal passages by keeping them in an unnatural state; some physicians, at least, ascribe the prevalence of colds in winter partly to this cause. Certain factories are kept artificially at constant humidity throughout the year on account of the effect upon the material (e.g. cotton) they use, and telephone (and other) companies sometimes find it profitable to keep their employees in a state of uniform comfort and efficiency by the same means. The cooling effect of the evaporation of moisture on one's skin in hot weather is an element of great comfort and serves to keep our body temperature constant. In the most accurate weighing, it has been found that an invisible film of water adheres to the weights when the humidity is high, and the errors thus introduced are easily detectable. Thus the humidity of the air is practically important, and records of it are taken daily by stations of the Weather Bureau.

Hygrometers. Instruments for measuring the humidity are called *hygrometers*. The most usual form is the *wet-and-dry-bulb* thermometer. If a thermometer bulb is wrapped with a bit of cotton dipping into water, it will be kept moist, and evaporation will cool the thermometer somewhat. How much lower it will read depends on the amount of movement of the air near the bulb, as well as on the humidity. If a standard wind, artificially produced, is always maintained when readings are to be made, the depression of the reading of the wet thermometer will vary in a regular and reproducible manner with the humidity of the air. The readings can be translated into real humidi-

ties by means of tables prepared from a series of calibration experiments, in which the humidity is found by passing a measured quantity of air through a drying agent, and weighing the water thus caught.

A rough form of hygrometer can be made from a long human hair, soaked in ether to remove traces of oil, and mounted so that small changes in length are magnified. It lengthens slightly with moisture, and is made to move a pointer over a calibrated scale.

Dew, fog and clouds. Air which contains a moderate amount of moisture will, if cooled, become saturated at a temperature known as ***the dew point,*** the numerical value of which can be judged from Fig. 14–6 if the moisture is given by its pressure, or from Table XVII if it is given by weight.

TABLE XVII

Weight of Water in 1 Cubic Meter of Saturated Air

Temp.	Wt. in grams	Temp.	Wt. in grams	Temp.	Wt. in grams
0° C.	4.84	15°	12.71	30°	30.04
5°	6.76	20°	17.12	35°	39.18
10°	9.33	25°	22.80	40°	50.70

If the air is cooled below the dew point, water will be condensed out of it in the form of fog, cloud, dew, or hoarfrost, depending on conditions. Dew can be observed on the outside of a polished metal can full of ice water, or a freezing mixture. Out-of-doors the dew forms first on the coolest objects; it forms on clear nights rather than on cloudy ones because, when the air is clear, the earth radiates away the heat which it has received during the day, and becomes cooler than the main body of the air itself. The coolest ground objects then form centers of condensation. This occurs most readily on small bodies, such as tips of grass blades, etc., or in free air, on particles of dust, as in the formation of fog.

When a mass of moist air rises, it is cooled by expansion and may reach a temperature so low that a cloud is formed. When this happens the condensation of water liberates a large quantity of latent heat, so that the temperature of the air ceases to fall, especially in the center of the ascending current. This, in turn, encourages the upward flow, and may actually make it very violent, carrying enormous billows of cloud to great heights to form the "thunderheads" (cumulus clouds) which are so common in sum-

mer. Their characteristic flat base discloses the level where the temperature reaches the dew point. Thunderstorms arise from the same state of affairs on a still larger scale, and usually depend for their existence on a layer of warm moist air near the ground.

Graphical representation of properties of water. If we should plot on one diagram all the conditions of pressure and temperature at which water can exist in ***two forms simultaneously*** in equilibrium, we should obtain a diagram like Fig. 14–7. The part AB is really a curve showing the boiling point at various pressures, or the pressure of saturated water vapor at different temperatures. The line AC is steep on account of the fact that large pressures produce only a slight lowering of the freezing point, an effect to be discussed in the next paragraph. Any point on this line gives a particular pressure and the temperature at which ice melts at this pressure. The line AD, all of which is below the usual freezing point, shows at any temperature the pressure of saturated water vapor when in equilibrium with ice. It is called the hoarfrost line. It is worth noting that even in ice the particles are in a state of agitation, and occasionally one flies off, or evaporates, and thus becomes a particle of water vapor. This evaporation produces a certain pressure of water vapor above the ice, and the largest possible pressure at any of these temperatures is the one given by the line AD. The point A is known as the triple point, when all three states of water can exist together in equilibrium, as already noted (p. 191).

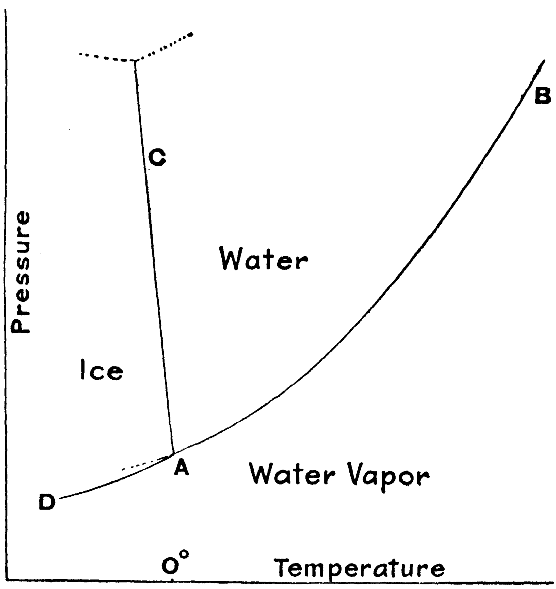


Fig. 14–7
The triple-point diagram for water

Pressure and melting point. The change in melting point with pressure is associated with the change in volume on melting. If

a substance shrinks on melting, as ice does, for instance, and we put pressure on it, we assist it in shrinking, and therefore make melting easier, or, in other words, the melting point is lowered. Conversely, if the substance expands on melting, the melting point is raised by pressure. The easiest way of showing this change is by what we shall call the ice-cutting experiment, in which a wire cuts its way through a piece of ice and yet leaves the ice whole. A heavy weight is attached to a fine wire (e.g. 2 kg. on a steel wire 0.5 mm. in diameter) which is thrown as a loop over a more or less rectangular block of ice, as shown in Fig. 14–8. The wire immediately begins to cut its way into the ice and eventually through it, but the ice then freezes solid again, as the wire goes through it. Immediately under the wire the pressure (the force per unit area) is very large, since the area is very small. Hence the ice there, which is at 0° C. to begin with, finds itself at a temperature which is above its melting point (since the melting point is lowered by the pressure). It therefore melts, abstracting the necessary latent heat from itself and the surrounding ice as it does so. Thus a little region around the wire is cooled to a temperature slightly below 0° C. Then the resulting water flows around the wire to the top, where it is relieved from pressure, finds itself at a temperature slightly below its freezing point, freezes again, and restores the latent heat which it had borrowed from the surroundings. Heat is thus continually being delivered at the top of the wire, flowing down through it, and being used at the bottom. The wire usually leaves a visible "track," since it alters the arrangement of the air bubbles which are always frozen into ordinary ice. Another example of this same phenomenon is furnished by the flowing of glaciers which is facilitated by the melting of ice under pressure.

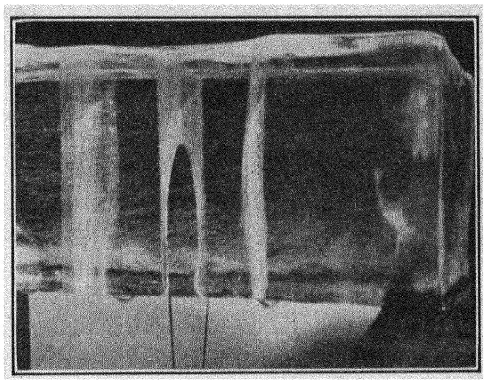


FIG. 14–8

Cutting ice by a loop of wire and yet leaving it solid. The wire is halfway through and the tracks can be seen which it made in going through several times before.

At extremely high pressures, Bridgman has found that there are

several forms of ice, the new ones all being denser than water. The melting points of some of these rise with the pressure, so that at room temperatures water freezes when the pressure reaches about 10,000 kg./cm.^{2}; and at 20,000 kg./cm.^{2} its melting point has reached 76° C.

Ice formation in rivers. Northern rivers show anomalous ice formation of considerable interest, occurring chiefly on clear, cold nights. In the most rapid places, the stream remains open all winter, and the surface water in contact with very cold air may be slightly undercooled and then mixed with the under water. Under these conditions a light spongy mass composed of fine, lightly-connected ice crystals forms under water, which is known as "frazil." This is carried downstream into quieter water, and under surface ice, where, if the heat conducted up from the body of the earth is insufficient to melt it, it may block the stream, and cause winter floods. In any case, its presence in water used for power purposes introduces serious difficulties.

On clear, cold nights the rocks over which open water flows become cooled by radiation of heat out through the water, so that the water freezes on them in a form known as "anchor ice." On clear days the heat from solar radiation is likely to warm the rocks to such an extent that this anchor ice breaks away and floats to the surface, sometimes bringing up bottom materials, and even rock fragments with it.

The temperature of the water varies very slightly in all such cases from 0° C., but changes of 0°.001 C. are significant in producing freezing or melting, and much interesting work in flood prevention and power development, involving very accurate temperature measurement, has been done on these forms of ice.[1]

Isothermal curves. The critical point. Any gas when near its condensing temperature ceases to follow Boyle's law exactly, and is then properly called a vapor. Its behavior is well shown by a set of observations embodied in a family of "isothermal curves," showing how the pressure changes with the volume at a constant temperature. Figure 14–9 gives such curves for carbon dioxide. At a fairly high temperature, the gas follows a curve like AB, which is nearly a true Boyle's law curve (pv = constant). At lower temperatures irregularities develop, and finally (as at 20° C.) a "curve" like $GHJK$ is obtained. Here the straight portion signifies that compression in that range diminishes the volume without increasing the pressure. This occurs only when the substance is a mixture of liquid and saturated vapor. Compressing such a mixture merely condenses part of the vapor into liquid, the

[1] The reader who wishes to know more of these matters is referred to "Ice Formation" by Dr. H. T. Barnes, 1906 (J. Wiley and Sons).

saturated vapor pressure (p. 194) remaining constant at the value appropriate to that particular temperature. Within the dome-shaped area in the figure, this condition prevails, the substance being partially liquefied. In the lower left part of the figure it is all liquid, and is hard to compress, as the steepness of the lines shows. In the lower right side, it is a thin vapor, acting much like a gas. If the temperature is high, no choice of pressure and

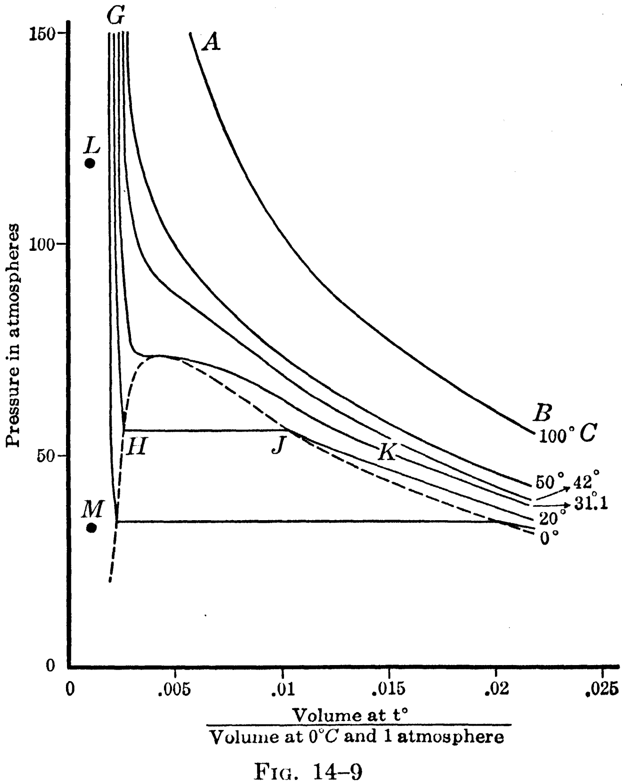


FIG. 14–9
Isothermal curves for carbon dioxide (after Andrews)

volume can be found which will make any part of the substance occur in the liquid form. It is evident that to liquefy it we must first cool it, at least to such a temperature that the isothermal curve corresponding to it will cut into the dome-shaped part of the diagram. The curve which just touches the tip of the "dome" corresponds to what is called the ***critical temperature.*** For carbon dioxide it is 31°.1 C.; its value for several gases is given in Table XVIII. With each critical temperature goes a corresponding critical pressure.

TABLE XVIII

	Critical temperature (degrees centigrade)	Critical pressure in atmospheres
Hydrogen	− 234	20
Air	− 140	39
Water	365	194.6
Carbon dioxide	31.1	73
Ammonia	130	115
Sulphur dioxide	155	79
Ether	197	36

A curious possibility is disclosed by these curves, that of passing by imperceptible degrees from a gas to a liquid without any break. Thus, let us imagine that we start with carbon dioxide in a cylinder, closed by a movable piston, in a condition specified by the point K, that is, a gas, well expanded and at a moderate temperature; then warm it to the state B; then compress it at constant temperature, following the curve BA, to a very high pressure; then cool it, passing leftwards to such a point as L; and finally diminish the pressure until it reaches the value at the point M. We then have circumnavigated the dome, and arrived at a liquid state at a moderate pressure and low temperature, and at no stage of the process has any change of state occurred. The chamber is full of gas to begin with, full of liquid at the end, and always full of a homogeneous material at all intermediate stages.

An interesting experiment that is best shown in the projection lantern can be done with a small, thick-walled glass tube filled half with liquid (preferably carbon dioxide) and half with vapor; that is, under such conditions of pressure and temperature as to correspond to a point within the dome. The pressure in this case is high, and the experiment should be performed with caution, with some protection between the tube and the bystanders, and with care not to allow the temperature to rise unduly. By surrounding the tube with water at a temperature slightly below 31° C. (for carbon dioxide), and adding some at a higher temperature, the temperature of the tube can be made to rise slowly, and the point on the diagram which represents the condition of the material will rise inside the dome to its upper boundary. When this boundary is crossed the surface of the liquid fades away and all distinction between liquid and vapor disappears. When the temperature is lowered again, the liquid appears first as fog, but soon settles to its usual form.

"Dry ice." Carbon dioxide is a curious substance. Under ordinary conditions it is a gas, but it can be bought as a liquid in steel cylinders under a pressure of 50 atmospheres or so. This liquid form cannot exist in the open air. One might suppose that it could be allowed to escape from the cylinder into a vessel and be caught there. When the experiment is tried a violent rush of very cold vapor occurs, which carries with it a fine white snow-like powder, that can be caught in a cloth bag. This is the solid form of carbon dioxide, and is found to be at a temperature of $-79°$ C. though the cylinder from which it comes remains at room temperature.

If a diagram like Fig. 14–7 (p. 198) is drawn for this substance, the triple point must lie at a temperature of $-57°$ C. and a pressure of 5.1 atmospheres. A pressure of one atmosphere cuts across the "hoarfrost line," just as a pressure of 2 mm. would do in the case of water. This means that at this pressure the substance can exist as a solid if cold enough, or as a vapor; but not as a liquid.

The reason why the vapor cools itself when it escapes is that latent heat must be given to the liquid in order to vaporize it, and since this is not supplied to any considerable extent from the surroundings, the liquid has to steal it from itself. Some of the liquid therefore freezes in order that the rest may evaporate.

The "dry ice" formed in this way is coming into scientific and commercial use as a very clean, safe, and odorless refrigerant. Its only drawback is its almost too low temperature; as it readily freezes mercury it is obviously too cold for some purposes, and bad "burns" can be produced by it if it is carelessly handled.

The liquefaction of gases. The methods of liquefying air and other gases can easily be understood in the light of the preceding paragraphs. If one wishes to liquefy air, for instance, one must first cool the air to a temperature below its critical point (for nitrogen $-146°$ C.) and then bring its pressure to a suitable value, so that its condition is represented by a point inside the dome of the diagram (like Fig. 14–7 drawn for air). If it is desired to have the liquid out in the open air, its pressure is fixed at that of the atmosphere, and the diagram will then show one temperature, and one only, at which the liquid and its vapor can exist together. This will, of course, be the boiling point of the liquid, which for air is anywhere from about $-191°$ C. when it is first made, to about $-184°$ C. when the nitrogen in it has boiled off, and the remaining liquid has

become almost pure oxygen. If liquid air is put under a higher pressure, the heat entering from outside will make its temperature higher; conversely, if it is boiled in a partial vacuum, it will cool itself to a point much below its usual temperature, the exact amount depending on how much heat is reaching it.

Apparatus for liquefying air. As liquid air is now made in quantities and has proved useful in many ways, the method of making it is of interest. Dry pure air is first compressed to a high pressure (some 2000 lbs./in.^{2}) by a pump (P in Fig. 14–10).

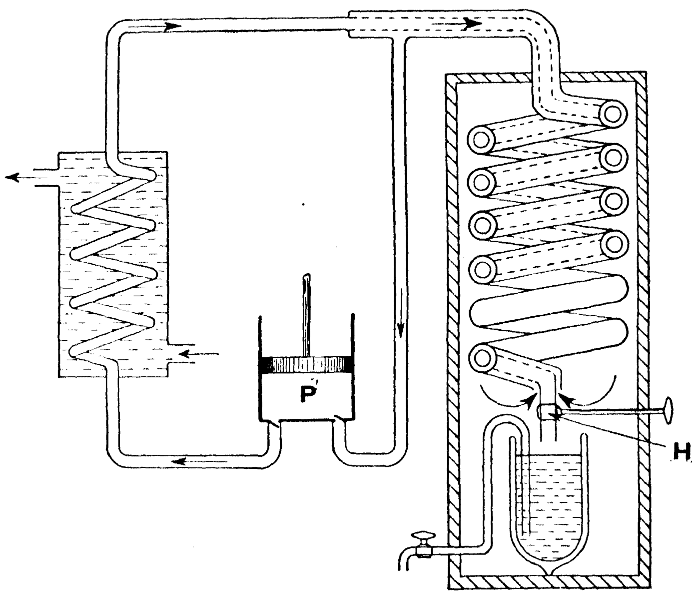


Fig. 14–10
Apparatus for making liquid air

This makes it very warm and the resulting heat is removed by passing the compressed air through pipes surrounded by a stream of cold water. It then passes into a long coiled tube (the "interchanger"), the action of which will be plain from the figure. The compressed air is driven through the inner of two tubes and escapes at the bottom through a small adjustable hole H (actually in a cock, controlled by a projecting handle). It is drawn back through the outer tube by the suction side of the double-acting pump. The expansion cools the air, largely on account of the Joule-Thomson effect (p. 168), which becomes especially large at low temperatures. Thus the incoming stream of air is cooled by the

outgoing one, and this cooling process is progressive, so that, if heat is prevented from getting in from outside, the expanding air eventually cools itself to a temperature below its boiling point, when drops of liquid air form by condensation, and fall into the vacuum vessel at the bottom, where it is kept until needed. The apparatus includes a valve (not shown) to admit more air to the pump and replenish the supply when liquefaction begins.

The liquefaction of other gases. ***Hydrogen*** is more difficult to liquefy. It is compressed, cooled by liquid air, and then allowed to expand in an interchanger, just as in the manufacture of liquid air. It must reach its critical temperature (− 234° C.) before beginning to liquefy, and in order to obtain the liquid at atmospheric pressure it must cool itself by evaporation to its boiling point (− 252° C.). By cooling compressed ***helium*** by means of liquid hydrogen and allowing it to expand, as before, it can be cooled to below its critical temperature (− 266° C.) and liquefied. Its boiling point is − 268°.5. By boiling liquid helium in a partial vacuum, with great precautions to prevent the entrance of heat, Kammerlingh Onnes, to whose experimental skill we owe most of these remarkable results, obtained liquid helium at a temperature only 0.82 degrees above the absolute zero. Later helium was solidified at a temperature of 3°.2 (absolute) and a pressure of 86 atmospheres. As there are no gases with lower boiling points, we may thus already have reached the lowest attainable temperature. The temperature was estimated from the vapor pressure, the curve of which can safely be followed by extrapolation down toward zero. All other thermometric methods become uncertain in this region.

The most remarkable discovery directly due to the use of these low temperatures was in regard to the practical disappearance of electrical resistance in certain materials (p. 340). To low temperature research in general must also be credited the discovery of three inert gases (neon, krypton, and xenon) existing in minute amounts in the air. Many curious physical phenomena have also been found to occur in this range. There is still much to be done at the lowest temperatures, hitherto made difficult by the exacting technique and the high cost of the necessary materials.

Experiments with liquid air. A few striking experiments might be mentioned which can be performed with a small quantity of liquid air. Before the liquid reaches the experimenter, its nitrogen has usually boiled off, leaving nearly pure oxygen. Some chemical experiments can be performed to show this,

such as thrusting the smouldering end of a stick into the liquid and observing the violent way in which it burns. Besides these, others of a purely physical character can be done. Mercury, alcohol, gasoline, etc. can be frozen. Lead can be turned into a brittle, elastic substance. A piece of beefsteak becomes like a rock which can be shattered into a mass of brittle fragments by a hammer blow. Rubber turns into a glass-like material. One experiment especially deserves a little description. Charcoal is a porous substance with a large area of surface (just as Norway possesses an abnormal length of coast line). On this account it shows to an unusual extent a phenomenon mentioned above, (p. 195); namely, an attraction between the solid surface and the particles of a gas. The gas is said to be "adsorbed," or "occluded" in the solid. The amount of this attraction depends on the kind of solid and the kind of gas. At ordinary temperatures charcoal will filter out certain gases from air; it was used for this purpose in gas masks during the World War. But at the temperature of liquid air charcoal attracts ordinary air to such an extent that it may act like a powerful pump and produce a fair vacuum. In Fig. 14–11 A is a tube filled with small lumps of hard charcoal, recently heated to drive off gases, and surrounded with liquid air. CDE is the mercury in a glass tube to which A is connected, E being open to the atmosphere. The connection is made by a rubber tube at the beginning of the experiment. The height CE will quickly rise to within a centimeter or so of the true barometric height. When associated with an ordinary air pump, a tube of charcoal in liquid air is of great service in producing high vacua. Its use for this purpose is now decreasing, following the invention of the more satisfactory diffusion pump (p. 141), but even with this pump liquid air is still commonly used to freeze out mercury vapor.

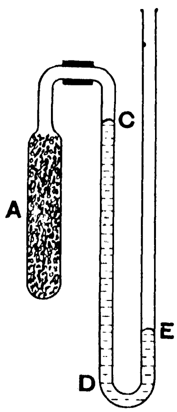


FIG. 14–11
The absorption of air by very cold charcoal

EXAMPLES

Example 1. One thousand grams of boiling water (100° C.) are poured into a heavy pitcher of 1000 grams weight and specific heat 0.2, containing 200 grams of ice, all at 0° C. What is the temperature when the mixture is complete?

The heat equation is formed on the assumption that the ice all melts, and the final temperature is $t°$ above zero. It gives:

$$\underset{\text{heat lost by water}}{1000(100 - t)} = \underset{\text{heat gained by pitcher}}{1000 \times 0.2 \times t} + \underset{\text{heat to melt ice}}{200 \times 80} + \underset{\text{heat to warm it}}{200 \times t}$$

from which $t = 60°$ C.

Example 2. Suppose that in the last problem there has been 1400 grams of ice in the pitcher. The heat equation, formed as before, yields

$$1000(100 - t) = 1000 \times 0.2 \times t + 1400 \times 80 + 1400 \times t$$

whence $t = -5°$ C.

This purports to say that by adding boiling water to ice one can lower the temperature of the ice. This absurdity comes from formulating the heat equation on a wrong assumption. We should have guessed that the final temperature is zero, and that the ice does not all melt. On this new basis we have

$$1000 \times 100 = M \times 80,$$

if M is the mass of the ice that melts; whence we find that we end with 2250 grams of water at 0° C. in the pitcher, with 150 grams of ice in it. Note that the pitcher does not now enter into the heat equation since it does not change in temperature.

One does not always know with what assumption to start in such problems, but a wrong one always leads to an absurdity. For instance, if in the first example, we had assumed that the final temperature was 0° C., and that the ice did not all melt, we should have formed the same type of heat equation as we finally did in Example 2, leading to the result that 1250 grams of ice would melt, which is absurd, since there are only 200 grams of ice in all.

Example 3. In one of the early experiments on the heat equivalent of work, the work was done by rubbing pieces of ice together. If 424 kg.-meters of work were expended in such an experiment, how much ice would be melted?

Assuming the heat equivalent to be 424 gr.-meters to the calorie, we see that 424 kg.-meters give 1000 calories, which would melt 1000/80 or 12.5 grams of ice.

Example 4. If live steam is passed through a hose into a pail containing 5000 grams of water at 10° C. (water equivalent of pail given as 200 grams) how much will have condensed by the time the temperature of the water has reached 40° C.?

The heat equation gives

$$\underset{\text{steam condensing}}{540 \times M} + \underset{\text{and cooling}}{M(100 - 40)} = \underset{\text{water and pail gaining}}{5200(40 - 10)}$$

whence

$$M = 26 \text{ grams.}$$

Books recommended:

W. J. Humphreys, "Physics of the Air," (McGraw-Hill).

M. Luckiesh, "Book of the Sky." Popular meteorology, 1922 (Dutton).

Tyndall, "Heat as a Mode of Motion," 1865 (Appleton).

Poynting and Thomson, "Heat," 1902 (Griffen and Co.).

PROBLEMS

1. Why is a burn produced by live steam worse than one caused by boiling water?

2. It is often said that hot-water pipes freeze during very cold weather more often than cold-water pipes. Can this be true? If the water in a hot-water pipe is frozen, why is it more likely to burst the pipe than water which has not been heated?

3. Why does a porous earthenware jar filled with water keep itself cooler than the air on a summer day? How would this cooling vary with the wind?

4. Why does food cook faster in a pressure cooker in which the steam pressure is considerably greater than that of the atmosphere?

5. The humidity in a room of 200 cu. m. volume is 20% at a temperature of 20° C. How much water must be evaporated in it to bring it up to 50%? (See Table XVII.)

6. Two hundred grams of lead shot at 100° C. are dropped into a dry cavity in a block of ice, where they are completely surrounded by ice at 0° C. What quantity of ice will be melted in the process?

7. A glass pitcher (weight 1000 grams, specific heat 0.2) contains 1300 grams of hot tea (thermally equal to water) all at 90° C. 1800 grams of ice at 0° C. is put in with the idea of making iced tea. What is the final state and temperature of the mixture?

8. Trace out just what happens when 1000 grams of ice at − 6° C. is mixed with 800 grams of water at 100° C. (Specific heat of ice = 0.5.)

9. Two hundred grams of ice are brought indoors on a day when the outside temperature is − 10° C., and quickly put into a vessel whose weight is 100 grams and specific heat 0.1, containing 290 grams of water at 30° C. Trace out exactly what happens when the mixture is made. (Specific heat of ice = 0.5.)

10. In the morning after a cold night a steam radiator made of iron (mass 50 kgm.; specific heat 0.1) is at − 7° C. When steam is turned on, how much must be condensed in the radiator so as to warm it up to 100° C., assuming no losses?

11. A specimen of a rare material weighing 5 grams is hung by a thread from a sensitive balance inside a chamber at 20° C. Steam is then passed into the chamber, bringing the temperature up to 100° C. Some of the steam condenses on the material, clinging to it. Its weight is thereby increased to 5.2 grams. Find the specific heat of the material.

12. A light dish contains 1000 grams of water. It is put on a stove where it rises in temperature at the rate of 10.8° per minute, when nearing the boiling point. How long will the water be in boiling away altogether, if the heat continues to enter at the same rate?

13. A pailful of snow is melted by passing steam into it. The pail weighs 2000 grams, and has a specific heat of 0.1. There are 2000 grams of snow in it, at − 10° C. to begin with; the final temperature of the mixture is + 20° C. How much steam was used? (Specific heat of ice or snow = 0.5.)

14. Ten grams of steam at 100° C. are delivered by means of a tube into a vessel containing 100 grams of ice and 200 grams of water at 0° C. The vessel weighs 50 grams, and has a specific heat of 0.1. What is the final state and temperature of the contents of the vessel?

15. A mass of molten lead is cooling down, and when near the point of solidification is falling in temperature at the rate of 10° C. per minute. How many calories per gram is it losing per minute? (Use a symbol for the weight of the lead.) If it continues to lose heat at this same rate, and takes 18 min. to solidify completely (remaining at constant temperature all this time), find the latent heat of fusion of the lead. (Assume the specific heat of molten lead to be 0.03.)

16. If in Fig. 14–1 (p. 187) the slope of the cooling curve BC at B is obtained, it represents the rate of cooling at the melting point, just as truly as the slope at A of the curve DA represents this same quantity. If these two slopes are different, and are both measured, show how this would lead to a value of the specific heat of the solid, if that of the liquid were known.

CHAPTER 15

HEAT ENGINES AND THERMODYNAMICS

Our modern mechanical civilization owes its recent rapid growth largely to the development of heat engines, machines which furnish work when heat is applied to them. This development has a long history, and recent triumphs in the invention and improvement of engines for ships, automobiles, airplanes, etc., are something of which we may well be proud. Still, the present situation is not entirely satisfactory. The fact that our natural sources of fuel are limited calls for careful conservation, and it is the business of the scientific man to devise ways of using this fuel most efficiently.

Fuels, heats of combustion. It is of interest to examine our natural sources of heat in order to find out which give us the largest returns. We derive heat most commonly from the combustion of coal, gasoline, wood, etc., though it is quite possible now to heat a house by electric power in regions where this is not expensive (e.g., near large waterfalls). Table XIX gives rough average values for the number of calories obtained by burning one gram of each substance.

TABLE XIX

Substance	Calories per gram
Coal	7,200
Fuel Oil	10,000
Gasoline	11,000
Alcohol	6,500
Hard wood	4,600
Gas	10,000
Hydrogen	34,000

The number of British thermal units per pound can be obtained from the figures given by multiplying each by 1.8. It must be

added that the values vary greatly with the quality of the substance, and that they refer to scientific tests in which ***all*** the heat that the substance gives was captured and measured. In practice this is not usual. An ordinary coal furnace may easily lose 30% of the heat obtainable from the fuel. In cooking on a gas stove, it is common practice to allow half the heat to escape. A beautiful wood fire glowing in the fireplace probably sends us much less than 10% of the heat liberated, the rest going up the chimney.

The relative cost of heating by means of different fuels is an interesting item which quite naturally varies greatly in different places. In most regions, coal gives the best return. Gas is often ten, and electricity a hundred times more expensive.

Early steam engines. The earliest "engine" was a curiosity and a toy. Hero of Alexandria (first century, A.D.) devised a globular vessel which when filled with water, boiling over a fire, would spin as a reaction from the stream of steam which was escaping from it. By 1690 a sort of steam pump had been made, consisting of a piston in a cylinder, which was occupied alternately by steam and by the water which was being pumped. In 1705 Newcomen arranged a separate boiler, so that steam entered a cylinder, and there pushed upon a piston, which was thus enabled to do useful external work. The steam had to be removed by condensation produced by a stream of cold water. By 1763 Watt [1] devised a separate condensing chamber, and in 1776 he invented the double-acting engine. In this form the steam is allowed to push the piston as far as it can in one direction, and then is automatically cut off by valves, and admitted instead to the far side of the piston, so as to force it back again. Since 1782 the cut-off has been arranged to occur early, so that the steam might cool itself by its own expansion, while it does work upon the piston. It is thus deprived of most of its available heat before it is discarded, with a corresponding increase in economy.

Modern steam engines. Steam engines are of two main classes, reciprocating engines and turbines. Both types have boilers in which the flames from the burning fuel may be drawn through pipes surrounded by the water, as in locomotive engines, or the water may pass through pipes about which the flame gases are made to circulate. In either case the hot gases give up nearly all

[1] James Watt (1736–1819), a Scotch instrument-maker and inventor, after whom the unit of power is named.

their heat to the water, turning part of it into steam at high pressure. In many modern engines of both classes this steam is superheated on its way to the place where it is used (e.g., the cylinder), so as to avoid the presence of any spattered water drops, and keep the steam "dry." In the ***reciprocating engine*** it then enters the cylinder by way of a valve chamber. In one arrangement (Fig. 15–1) the cylinder has two openings which can interchange their functions. One serves at one moment to admit steam to the left side of the piston (as shown), while the other allows the spent steam on the right side to escape; then, through the action of sliding valves, the second becomes an entrance instead of an exit, and the first

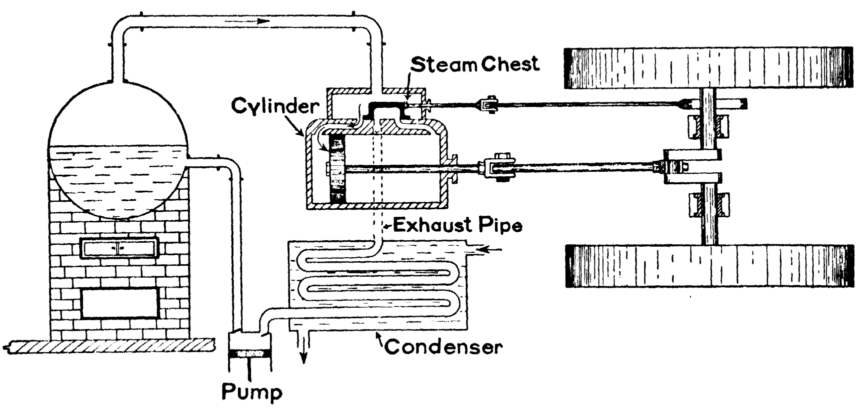


Fig. 15–1

becomes connected with the exhaust pipe. Thus the piston is driven by the steam pressure in each direction.

In various ways some of the heat goes astray in the steam engine (as well as in the boiler where the fuel is burned), so that we get no work out of it. In locomotives, for instance, the spent steam is discarded through the exhaust pipe, carrying off a considerable amount of heat with it. The walls of the cylinder absorb heat, especially if the steam is at high pressure when it enters, and is used in only one cylinder. The entering steam must then be at a high temperature, since high pressure and high temperature go together, according to the curve (p. 194) of saturated vapor pressure. Hence this steam gives up some of its heat to the cylinder in the first part of the stroke, and keeps it so hot that the exhaust steam cannot fall to so low a temperature as it should when it expands. It is more efficient to provide for several stages of

Figure 15–2 shows how this is done. Figure 15–3 gives some idea of the size and general arrangement of a real turbine.

Turbines are most efficient at high speeds. They are often mounted on the same shaft with electrical generators. They must be geared down if they are to move ordinary machinery. When large ships are driven by turbines, these engines are often used to create electrical energy, and this is then transmitted to the slower-moving motors which drive the propellers.

The internal combustion engine. The gas engine, of which the gasoline engine is one form, is quite different. Here there is no

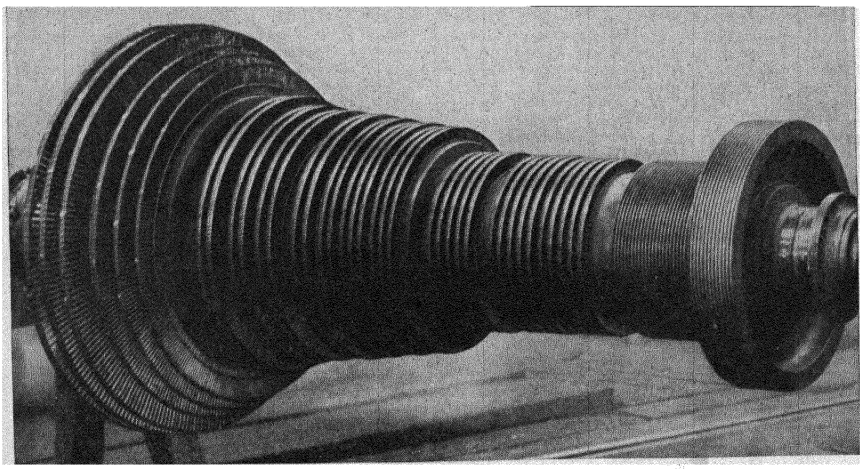


Fig. 15–3

The rotor of a steam turbine, showing (at least in the larger discs) the vanes against which the steam strikes. Stationary diaphragms, fastened to the cover of the rotor, fit in between the rotating discs. The high-pressure steam enters on the right and expands as it goes to the left.

boiler; the fire is within the cylinder. The fuel is a mixture of an inflammable gas, or vapor, and air, which is usually made in a "carburetor" and drawn from there into the cylinder through a valve (Fig. 15–4*a*). This valve then shuts, the piston compresses the charge (*b*), and when it is most compressed, the mixture is fired by an electric spark, which is arranged to occur at a spark plug by means of suitable electrical apparatus. The resulting heat expands the gases within the cylinder (*c*), giving a violent push to the piston at that stage of the action. At this stage, the cylinder contains the waste products of combustion, which are removed, in the usual type of automobile engine, by the "scavenging

stroke" of the piston (*d*). This drives them into the exhaust pipe, through the exhaust valve, which opens at that time only. In this type of engine the piston has to go back and forth twice for each impulse of power, and these four motions have given it the name of the "four-stroke" or "four-cycle" engine. "Two-cycle" engines are common in motor boats; toward the end of their "power stroke" their exhaust opens and the products of combustion blow out to a certain extent while the fresh charge is entering. The scavenging action is incomplete, and such engines are less efficient, though lighter. The four-cycle gas engine is usually supplied with at least four cylinders, all working with power strokes timed successively, to make the same shaft revolve. For each

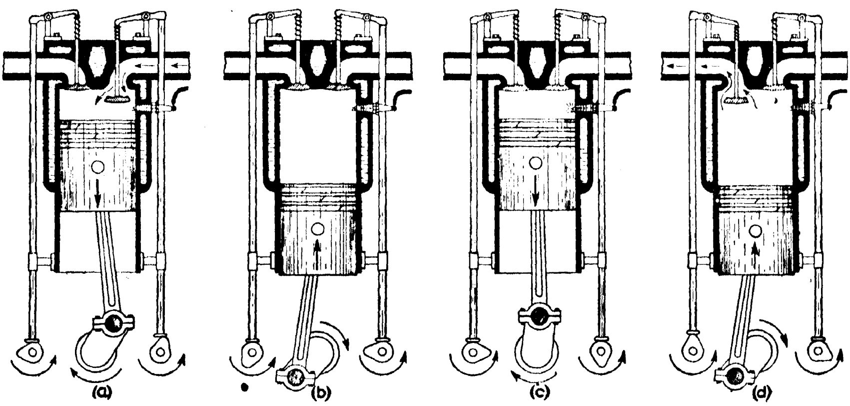


FIG. 15–4
The four stages in the action of one cylinder in an automobile engine

cylinder the engine makes two complete revolutions between power strokes; when there are four cylinders it turns only half a revolution between explosions; with six, one third. Thus a modern six-cylinder automobile engine furnishes "power strokes" so frequently that the car when in high gear has to move only a foot or two between them; in low gear only a few inches. Thus the driving impulses become so nearly continuous that the action is very smooth.

The ***human body*** may also be regarded as a sort of internal combustion engine. Fuel is consumed, heat is lost by radiation, evaporation, etc., and a fair proportion (15% or so) of the energy supplied may be turned into mechanical work. Interesting research on this subject has been carried out on individuals isolated

in a "calorimeter" at the Carnegie Institution Nutrition Laboratory in Boston.

The Diesel engine. The Diesel engine is an interesting form of internal combustion engine because of its high efficiency. In this type (Fig. 15–5) air is drawn into the cylinder and compressed to a very high pressure (up to 500 lbs./sq. in.) so that it is made very hot, and has no time to cool off to the cylinder walls before a fine spray of oil is spurted into it. The oil takes fire from the heat of the air itself. The burning of the oil gives a power stroke,

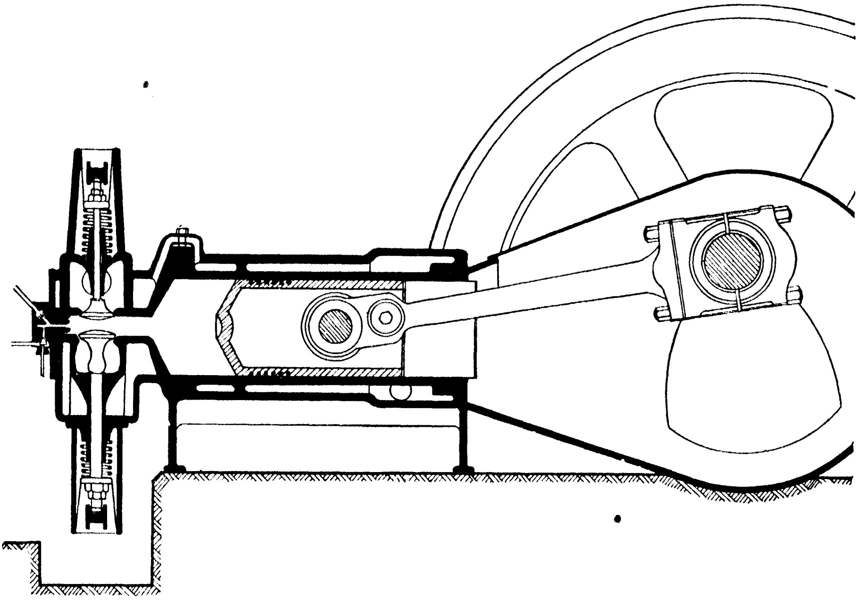


FIG. 15-5

A section of a horizontal Diesel engine. On the extreme left is the fuel spray; above and below it are the two valves, opening into the small space into which the piston in its leftward motion compresses the air.

which is followed by a scavenging stroke, completing the cycle. The combustion is relatively slow, so that as the piston is driven back the mixture expands, doing work throughout the expansion. A low-grade oil may be used as fuel. This type of engine has the advantage of needing no carburetor or ignition system. On the other hand it is heavy, expensive to build because of the high pressures, and it cannot be started without a small auxiliary source of power to act as compressor.

Efficiency of engines. The efficiency of an engine is the ratio of the amount of heat actually turned into mechanical work to

the amount given by the fuel consumed. In good locomotive engines it now reaches about 10%, in triple-expansion marine engines about 20%, in steam turbines 25%, in a good automobile engine 25%, and in the best Diesel engines 34%. To make the comparison fair, the boiler losses are, of course, included in the figures for steam engines.

These figures look as though they ought to be misprints. It seems at first sight absurd that the efforts of skilled engineers for more than a century have culminated in results representing so shocking a waste of our natural resources. Table XX shows where some of the wasted heat goes in a few sample types of engines. It is interesting, in order to see why we can do no better, and to get a better understanding of the nature of this problem, to enter briefly into the subject of thermodynamics, the science dealing especially with the relations between heat and work.

TABLE XX

Losses in Engines

Steam engine		Automobile	
Work done	9%	Work done	21%
Rejected to condenser	57	Into cooling water	36
Up chimney	22	Out the exhaust	35
Boiler losses, etc.	12	Friction	8

The first law of thermodynamics. The first law of thermodynamics is nothing but the principle of the conservation of energy. Any mechanical change that occurs in an isolated system, or body, (i.e., one not acted on by outside forces) is subject to the general law that the energy of the system remains constant, heat being considered as one form of energy, and being measured and counted in with the rest. The fact that perpetual motion machines will not run is one reason why we have confidence in the truth of this law.

The second law of thermodynamics. The second law of thermodynamics is to the effect that heat cannot ***of itself*** pass from a body of low temperature to one of high. This sounds somewhat like a back-handed statement of what we mean by temperature; but there is more in it than that. We may imagine cases in which the law appears not to hold. For instance, a compressed gas in one cylinder may by expansion succeed in pushing a piston and by this means compress another gas in another cylinder, thus raising

the temperature of the second gas while that of the first goes down. This appears to be a flow of heat from a cold body to a warm one. It is, however, a temporary flow of energy, and takes place at the expense of the potential energy stored at the beginning in the compressed gas. It cannot go on unless the original store of energy is replenished by an outside agency; and this is not a flow of heat "by itself." Other apparent violations of the law can likewise be explained away. We now accept it as of equal validity with the statement that water cannot of itself run uphill.

Carnot's engine.[1] The two principles just stated govern the action of heat engines. The theory which follows is based on them, and was developed by S. Carnot [2] (1824). He imagined an engine which is too nearly perfect ever to become a reality, and then proved that this particular type would be the most efficient engine conceivable, if it could only be made. Even though we shall never see one, we are interested in it, as it shows the limit to the success attainable by our efforts, and points to the most hopeful directions for improvement.

Carnot imagined an engine consisting of one cylinder (and other parts which we shall not consider), with perfectly non-conducting walls and piston, and a base which could be altered at will from a perfect conductor to a perfect non-conductor. He further imagined two large reservoirs of heat, one at temperature T, and the other at a higher temperature T'; they must be large so that nothing that we do to them will sensibly alter their temperatures. He supposed the cylinder to be filled with a gas — a perfect gas, that is, one following the simple gas laws. The engine thus far reminds one distantly of a hot-air engine, such as one occasionally sees in toy sizes, which works by the expansion of air with heat. The gas is called the "working substance."

Isothermal and adiabatic changes in a gas. Before working out the cycle of operations of Carnot's engine, we must consider once more the behavior of a perfect gas. When it is kept at constant temperature, we have seen (p. 118) that its pressure and volume

[1] In a short course this discussion (four paragraphs) might be omitted.

[2] Sadi N. L. Carnot (1796–1832); French physicist; original and profound as a thinker, but called by his duties as an army officer to occupy himself with other work. His notebooks contain suggestions for fundamental experiments on the nature of heat, which were actually performed by others, but not until some twenty years later.

follow a family of Boyle's law curves (isothermals) shown in Fig. 15–6 by continuous lines. If, however, a gas is treated in this way, heat must always be added to it when it expands to keep it from falling in temperature. There is another simple way in which we can imagine the condition of a gas to be changed; it might be compressed or expanded without adding to it or taking from it any heat at all. Such a change is called an ***adiabatic change.*** The dotted curves in Fig. 15–6 show the change in pressure and volume under these circumstances, and are called adiabatic curves. Their equation is shown in books on thermodynamics to be $pv^{\gamma} = c$, where c is a constant, and γ ("gamma") is the ratio of the two specific heats of the gas. This equation can be used to get their exact form. We see at once that such curves must be steeper than isothermal curves, because, when a gas is compressed without taking any heat away from it, its temperature must increase, and this means that the curve must rise faster as it goes to the left than the isothermal curve does.

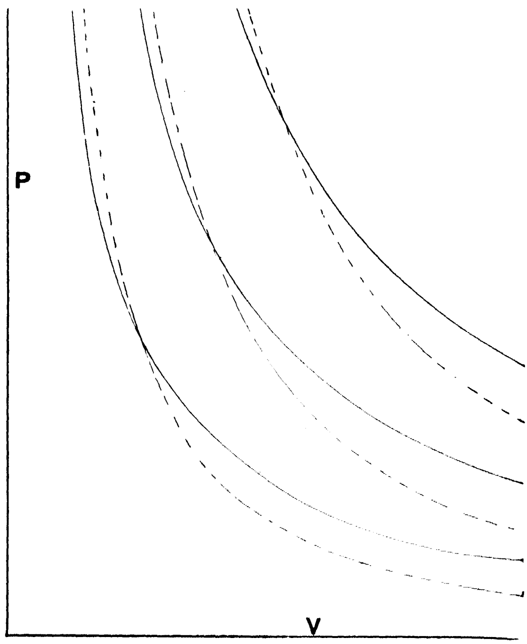


FIG. 15–6
Isothermal and adiabatic curves

Carnot's cycle. Now, to follow out the operation of Carnot's engine, let us imagine the gas in the cylinder to be at first rather compressed, and at the absolute temperature T', as represented by the point A in Fig. 15–8. The cylinder is supposed to be fitted at this stage (a) Fig. 15–7 with its perfectly conducting base, and placed in perfect contact with the reservoir of heat at the temperature T'. Now let the gas expand and push the piston back, doing external work. If this takes place slowly, the temperature never falls perceptibly during the process, as the contact with

the heat reservoir is perfect. A quantity of heat H' flows into the gas during this process so that it remains throughout at the temperature T' and the curve AB (Fig. 15-8) representing the change is part of an isothermal curve. Then, when the gas reaches the stage drawn in (*b*), Fig. 15-7, and represented by B, Fig. 15-8, the reservoir is removed, the non-conducting base substituted, and the gas is allowed to expand further, to the state (*c*), Fig. 15-7, or C, Fig. 15-8, doing more work during this change. This will be an adiabatic change, and it is continued just long enough so that the temperature falls to T. Then, as the third step in the cycle, the non-conducting base is removed and the perfectly conducting one replaced, with the reservoir at temperature T below it. The gas is compressed at temperature T, driving a quantity of heat H into the reservoir, and doing work upon the gas from outside. This stage is stopped at a point D, so situated that a return can be made by an adiabatic compression DA, (Fig. 15-8) back to the starting point, with the non-conducting base in use, more work being done upon the gas during this compression.

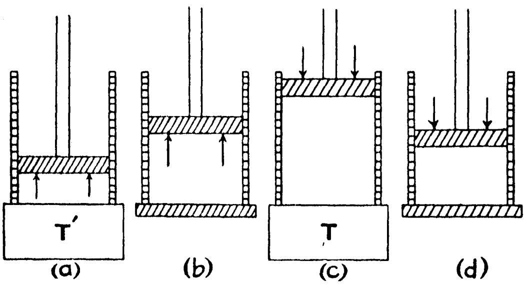


Fig. 15-7
Carnot's cycle

During the complete cycle the gas was supplied with a quantity of heat H' when it was at the temperature T', and gave out a quantity H when it was at the lower temperature. It supplied external work in the steps AB and BC, and took in external work (i.e., had work done on it) in the other two stages. Let us consider how much work was involved in all. Isolate a very small part of the change, from the state L to the state M, in the change AB (Fig. 15-9). The work associated with this change is the pres-

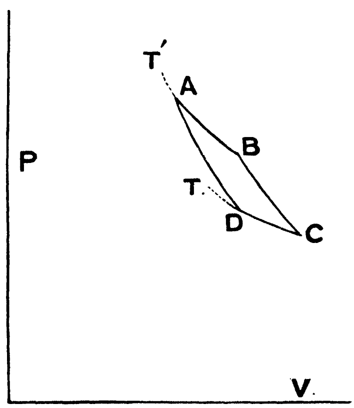


Fig. 15-8

sure at either L or M, (since they are extremely near together) multiplied by the small change of volume [1] represented by RS; in other words the work is represented by the area $LRSM$. Extending this process throughout the change AB, we see that the entire work done in this stage is the area under AB in Fig. 15–9. If, in Fig. 15–8, all such areas are drawn (they are omitted to save confusion) and account is kept as to whether the work is done ***on*** or ***by*** the gas, it is clear that the area $ABCD$ represents the difference between the work done by the gas and the work done on it in the complete cycle; or, this area represents the net work done by the gas. By the first law of thermodynamics, this work (measured in heat units) must be equal to $H' - H$, or the heat used up in the process.

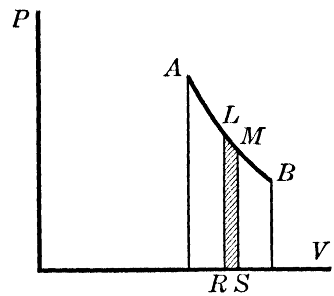


FIG. 15–9

Reversibility of Carnot's engine. Carnot's engine possesses the unique property of being reversible. This means not only that it could be forced to go through its cycle backward, but that, if it were, each heat transfer would occur in exactly the same amount as before, though in an opposite direction. If we think of an ordinary engine, in which the heat of a flame takes the place of Carnot's hot reservoir, it is evident that by running the engine backward, we could not put all the heat back into the flame, and the heat transfers would be quite different. In Carnot's engine, we could make the engine go through the cycle in the order $ADCB$, but we should then be doing more work upon the gas than we should get out of it, and we should be taking a quantity of heat H out of the lower reservoir, and pumping a greater quantity H' into the upper one.

This reversibility makes this type of engine the most efficient one conceivable. The efficiency of an engine is the ratio of the heat turned into mechanical work to the total heat put into the engine; or,

$$E = (H' - H)/H'.$$

[1] If a gas is contained in a cylinder at a pressure P, and the piston, which has an area A, is pushed in a small distance d, the total force exerted on the piston is $P \times A$, and the work done in exerting this force through the distance d is $P \times A \times d$. But $A \times d$ is the change in volume; hence the pressure times the change in volume equals the work done.

Now, if there were any sort of engine working between the same temperatures T' and T which was more efficient than the Carnot engine, one such could be coupled to a Carnot engine so as to run it backward. Let the piston strokes be adjusted so that the amount of work done by the new engine is just equal to the amount needed to run the Carnot engine. If the work done is the same, the heat used up must be the same, according to the heat equivalent of work. Then, if in the new engine a quantity of heat H_1' is drawn from the hotter reservoir, and H_1 is given to the cooler one, $H_1' - H_1$ is the heat used up in driving the Carnot engine backward, and this must be equal to the heat $H' - H$ which is delivered by the Carnot engine to the high-temperature reservoir, as a result of each cycle. But, if the efficiencies are unequal, as supposed,

$$(H_1' - H_1)/H_1' > (H' - H)/H' \quad \text{and} \quad H_1' - H_1 = H' - H.$$

Therefore

$$1/H_1' > 1/H', \quad \text{or} \quad H' > H_1'; \quad \text{and finally} \quad H > H_1.$$

This means, when written out in words, that the Carnot engine draws more heat from the lower reservoir (H) than the other puts into it (H_1), and puts more into the upper reservoir (H') than the other takes from it (H_1'). Hence, as a result of the combined action of the two engines, heat passes from the cold reservoir to the hot one without any work being done on the pair of engines, forcing it to flow in this unnatural direction. This is contrary to the second law of thermodynamics. Hence this cannot happen, and therefore no engine can be more efficient than the Carnot engine.

Theoretical engine efficiencies. As the thermodynamic treatment of engines is too long a subject for us to follow in detail, we shall shorten the discussion by accepting without proof one important conclusion. It can be shown that the efficiency of a Carnot engine, already given as $(H' - H)/H'$, can be written as a similar expression in terms of temperatures, if a new scale of temperatures is adopted. This is ***Kelvin's thermodynamic scale,*** which arose from his study of Carnot's engine, and was devised to rid our ordinary scale of the peculiarities characteristic of any one real substance. Kelvin's scale is independent of the substance, while the temperatures which we have so far used have been

derived from a thermometer filled with some sort of gas. Real gases fail to follow the gas laws with precision, and differ in this respect among themselves. Thus no two gases give exactly the same scale. An ideal gas, that is, one following these laws exactly and possessing no attractions among its particles, would be free from all defects, and it can be shown that such a gas would give Kelvin's scale precisely, when it was used in a gas thermometer. The deviations between Kelvin's scale and that yielded by a hydrogen thermometer are less than a hundredth of a degree in the ordinary temperature range, and rise only to one-tenth at the extremes. Such small errors are seldom important. Hence for measurements of ordinary precision there is no practical difference between Kelvin's scale and the hydrogen (or helium) gas scale that we have so far been using.

The efficiency of a Carnot engine, expressed in Kelvin's absolute temperature scale, is

$$E = (T' - T)/T'.$$

In honor of Kelvin (see footnote p. 168) we commonly designate temperatures on his absolute scale (which is also the ideal gas scale) as ° K., and this symbol will be used henceforth. This new formula for the efficiency enables us to calculate it as soon as the working temperatures of the engine are given, and thus we can find the ***upper limit to the efficiency of any engine whatever,*** working between these temperatures, since no engine that we can ever devise can be quite so good as the Carnot engine.

Thus, if the working substance of a real engine (steam, perhaps) is put through a cycle whose upper temperature is 200° C. (corresponding to a steam pressure of 225 lbs./in.^{2}) and lower temperature 100° C., these temperatures are $T' = 473°$ K. and $T = 373°$ K.; whence $E = (473 - 373)/473 = 21\%$. That is, more than three-quarters of the heat put into the engine is thrown away, even though the engine is perfect. Thus we see that low engine efficiencies are inevitable unless the upper temperature is raised, or the lower one made still lower, or both.

An actual steam plant involves other losses also; the main one being a boiler loss of 20% or more. In any figures for the efficiency this should be included when comparisons are made with internal combustion engines.

In condensing steam engines the pressure is very low in the

condenser (1/15 that of the atmosphere in a reciprocating engine, and four times less in a turbine) and thus quite low condensing temperatures are possible, reaching to about 25° C. As this is a limit, we must look for further improvements in the efficiency in the other direction, namely by raising the upper temperature. In a large plant in Massachusetts the upper temperature is 310° C., or 590° F., and the pressure has the enormous value of 1400 lbs./in.^{2} With a lower limit of 30° C., the Carnot engine efficiency becomes 65% between these temperatures. By superheating the steam in such a plant an efficiency of over 75% has actually

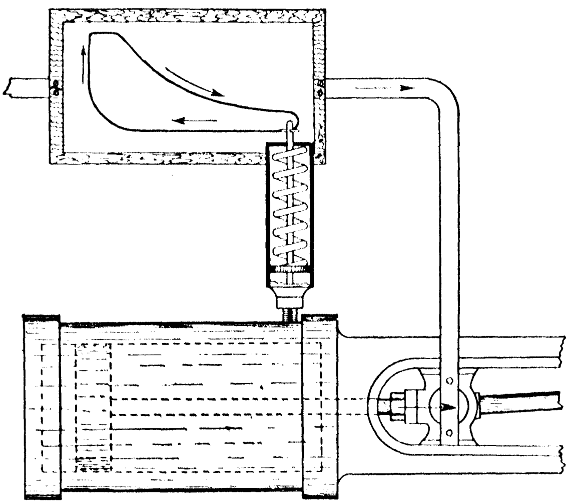


Fig. 15–10
An indicator diagram

been reached. It must be remembered, however, that these figures refer to the engine only. If it drives an electrical generator there is a small loss there, and the boiler always involves a considerable loss also. If all the losses are included, as they should be in making the figures for steam engines comparable with those for internal combustion engines, the result is that the best steam plants hardly reach the efficiency of Diesel engines.

An interesting recent experiment is the use of mercury vapor as a working substance instead of steam. A mercury turbine is in use in Hartford, Conn., which has an upper temperature of 458° C. (856° F.) and a pressure of 70 lbs./in.^{2} The lower temperature is high, but the heat derived from the condensing mercury is used

to create steam, which in its turn runs a steam turbine. The mercury turbine has an efficiency of about 30%; the two together, 25%. The cost of the necessary mercury (15 tons) is a considerable item. As mercury vapor is poisonous, the leakage problem is both important and at these temperatures difficult.

Gas engines carry the working substance to very high temperatures, momentarily at least, by having the explosion within the cylinder; hence their relatively high efficiency. It would be still higher if it were not for the rapid loss of heat to the walls of the cylinder.

Indicator diagrams. In reciprocating steam engines, a device is commonly attached which produces what is called an ***indicator***

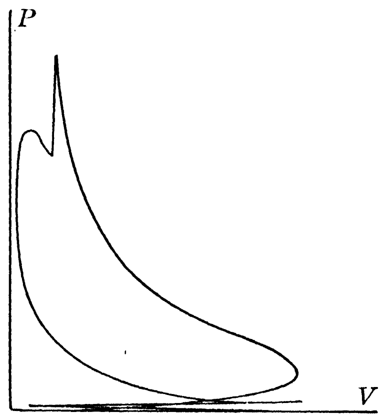


FIG. 15–11
Indicator diagram of a gasoline engine when knocking

diagram, very like the diagram of the Carnot cycle, except in shape. As the piston moves back and forth it carries a sheet of paper back and forth under a pen; as the pressure changes, it causes the pen to be moved up or down at the same time as shown in Fig. 15–10. The result is that a loop is traced by the pen on the paper. The area of the loop is proportional to the work done in the cycle, (see footnote, p. 221). The abscissæ as drawn are not actual volumes, but are proportional to them. The diagram is used especially to arrange the most efficient timing of the valves. When the area within the curve is as great as possible the engine is doing its best.

Somewhat similar diagrams are useful in studying the action of internal combustion engines. For instance, if a gasoline engine

is behaving normally it gives a smooth indicator diagram, but if it begins to "knock," a diagram like that in Fig. 15–11 is obtained. The knocking is due to an explosion (or to more than one) which is so sudden that the piston cannot move much in response to it.

PROBLEMS

1. A "shooting star" is a fragment of rock, or ore, traveling at a speed greater than that of any bullet, perhaps 30 km. per second. How does its kinetic energy compare in amount with the heat energy that can be obtained by burning an equal weight of coal?

2. Compare the cost of heating a house by coal and by electricity. Assume the cost of the latter to be 4 cents a kilowatt hour, and none of it to be wasted; but make an allowance of 50% for the heat of the coal which goes up the chimney. (Take the cost of coal as $15.00 a ton, roughly 900,000 grams.)

3. Find how many joules of work can be obtained from 2.65 kg. (about 1 gallon) of gasoline, when used in an engine of 25% efficiency. Convert this answer into kilowatt hours and horse-power hours. (746 watts = 1 H.P.)

4. If an automobile engine working at the rate of 30 H.P. uses 3.3 gallons (1 gal. = 2.65 kg.) of gasoline per hour, what is its efficiency?

5. How many tons of coal does a steam plant of 15% efficiency require per hour if it is to develop 1000 H.P.? (Take 1 ton as 900,000 grams.)

6. Find the power (in kilowatts) which an engine develops, given the following data: Piston area, 400 cm.^{2}; length of stroke, 70 cm.; number of strokes per second, 2; average pressure on piston during stroke 10 kg./cm.^{2}

7. A steam engine working between the temperatures of 250° C. and 50° C. has an efficiency of 25%. How does this compare with the efficiency of a Carnot engine working between the same temperatures?

8. How much would the efficiency of a Carnot engine be improved if its lower temperature was changed from 80° to 40° C. while the upper temperature remained at 300° C.?

9. The efficiency of an engine is 25%, and of its boiler 75%. How many H.P. will it develop per pound of coal used per hour? (1 lb. = 454 grams.)

10. Each of the two cylinders of a locomotive has a diameter of 1 ft., a stroke of 2 ft., and an average pressure of 200 lbs./in.^{2} At 5 strokes per second find the horse-power derived by the engine from the steam.

SOUND

CHAPTER 16

VIBRATIONS

Vibratory motion. There is one type of motion which we have not yet considered, though it is perhaps the very commonest of all motions, occurring more frequently even than that of falling bodies. This is vibratory motion. We find it in all sources of sound, in waves, in electrical oscillations and in the vibrations of all sorts of mechanical structures, such as automobiles, buildings, bridges, and even in the body of the earth itself. Since this type of motion is so universal in its occurrence, it is worth while to consider its nature and the causes that produce it.

Nearly every sort of material is more or less elastic and will resist any attempt to deform it, whether it takes the form of a push, a pull, a compression, or a twist. If a door slams or a heavy truck rattles by on the street, the entire building is shaken and the vibrations which are produced are felt not only near the source of the disturbance, but travel almost instantaneously through the whole building. When an earthquake occurs, masses of rock may be moved about suddenly, but the tremors they produce reach all over the world and may be recorded, in the form of vibrations, by the delicate instruments (p. 253) constructed for this purpose. These examples indicate that disturbances may give rise not only to ***vibrations*** at their source, but also to ***waves*** which pass these vibrations on to distant points. The latter are considered more in detail in Chapter 17.

Cause of simple periodic vibrations. In all common cases the cause of a vibration is some sort of a displacement which is resisted by an elastic force arising in the body itself. We have already seen

that an elastic body follows Hooke's law (p. 122); that is, the displacement gives rise to an elastic force resisting it, and this opposing ***force is proportional to the displacement.*** For example, consider a straight thin strip of wood or steel, clamped firmly in a vise at one end and bent to one side (Fig. 16–1) by a force acting at A, as indicated. If this force is suddenly withdrawn, the strip swings back and forth, the tip waving between A and B with a rapidity which depends on its length, mass and stiffness. At every position of the strip between A and B the elastic force tending to straighten the strip is proportional to the displacement; that is, if the end of the strip is at a certain instant situated halfway out between its mid-position and the extreme position A, the elastic force moving the strip is then just half as much as it was at A. The sort of motion that ensues when ***the force is strictly proportional to the displacement*** is called ***simple periodic motion.*** If the proportion is not a simple one, a more complex type of periodic motion results. This happens very commonly (for example, in our own hearing mechanism) and is interesting, but we shall defer consideration of it until the simpler cases have been examined.

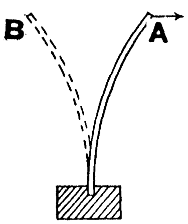


Fig. 16–1

The vibrations in any case arise through the combined effects of ***force*** and ***inertia.*** When the body is displaced or deformed, the restoring force generates a motion and the inertia carries this motion through, past the place of equilibrium, to an opposite displacement. For example, the strip in Fig. 16–1, when displaced to A and let go, gains speed while it is straightening itself out, and the inertia then carries it over to B, instead of allowing it to stop midway. If the strip were a feather, its inertia would make it act in the same way in a vacuum, but in air the friction would be so great that the motion would last only a short time.

Types of vibrations. It is possible for a body to vibrate in several different ways. For instance, the strip in Fig. 16–1 might have such a shape that it was half as thick as it was wide. In this case it would be easy to make it bend in the direction of its thickness, but it could also be bent in the direction of its width, and since it is stiffer in this direction, the elastic forces resisting this sort of bending would be greater, and the resulting motion, when the strip was released, more rapid. We might also twist the free

end of this strip without moving it to one side or the other. If this twist were suddenly relieved, the strip would vibrate with a twisting motion. Furthermore, it could be struck by a hammer, the blow being directed along its length, so as to compress the strip, whereupon oscillations of a different sort would occur, though probably we could not observe them very easily in this case.

Thus this strip can readily be made to vibrate in four different ways. The first three involve motion of the particles of the strip in a direction perpendicular to its length, whence they may be called ***transverse vibrations.*** The compressional disturbance involves motion of the strip particles along the strip; hence ***longitudinal vibrations*** arise.

Both sorts of vibrations may readily be shown in a steel rod, preferably of at least an inch in diameter and two feet in length. If this rod is resting horizontally on two thin wooden supports placed at points about a quarter of its length from each end, a blow in the middle of the rod will give rise to transverse vibrations furnishing a musical note of rather low pitch. If the same rod is then held by the fingers at its middle point and struck a longitudinal blow on its end by a hard hammer, it will emit a persistent note of very high pitch, which may be so high as to be inaudible unless the rod is long. This note is caused by longitudinal vibrations.

Another simple example is furnished by a spring brass wire wound in the form of a helix, such as is readily obtained by releasing the wire from a small spool. (Soft brass wire will not do.) If this is hung from the ceiling and reaches nearly to the floor, the experimenter may easily originate transverse vibrations in it, and (by compressing several turns and suddenly releasing them without transverse motion) longitudinal vibrations also. The latter are quicker, as in the case of the rod, and not so easy to follow. If bits of white string are tied to the wire a torsional vibration may readily be followed, which may be started by giving the bottom of the coil a quick twist.

Simple periodic motion. In order to get a better acquaintance with simple periodic motion, it is well to consider the oscillations of a common pendulum, which can easily be shown (p. 231) to occur according to the same law followed by the vibrations of elastic bodies. In Fig. 16-2, A is a ball hung on a string and swinging back and forth through a distance AB (which must be assumed to be small compared with the length of the pendulum). Its position of rest is at C. The distance AC or BC is known as the ***amplitude,*** which is half the extreme width of the swing. If the ball is at one instant passing through the point D, the distance CD is (very naturally) called the ***displacement.*** The time of one complete vibration, over and back, is called the ***period,*** or the

periodic time. The reciprocal of this is the ***frequency,*** or the number of vibrations per second, a term commonly used in connection with rapid oscillations. The part of the oscillation already performed, counting from some position arbitrarily chosen as a starting point, is called the ***phase.*** The phase may be measured as a fraction of a whole period, or in terms of angles, as shown below.

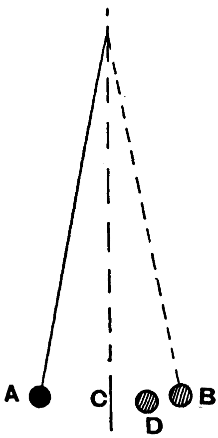


Fig. 16–2

Two similar pendulums vibrating together and moving alike at every instant are said to be in the same phase; if their motions are always opposed, they are said to be in opposite phases. All these technical terms may be used in describing transverse, longitudinal or any other type of vibratory motion.

Simple periodic motion is different from any type hitherto considered. The velocity is not constant, and the acceleration also is always changing. An easy way of discovering the peculiar properties of this type of motion is to derive it as the projection of uniform circular motion. In order to do this, let us suppose that a particle P is revolving at uniform speed in a circle, carrying out a type of motion already studied (p. 70). If M (Fig. 16–3) is the foot of a perpendicular drawn from P (wherever it is) on the diameter AB, the point M will oscillate back and forth with simple periodic motion between A and B, like the pendulum, as P goes uniformly around. The point M may be called the projection of the point P on the diameter AB. The circle is sometimes called the "circle of reference."

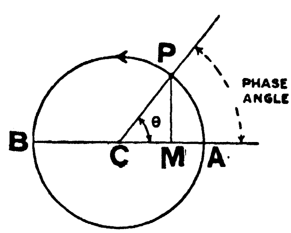


Fig. 16–3

We may use the pendulum in Fig. 16–2 to illustrate this sort of motion, if we give it such a start that the ball moves in a horizontal circle. Then we can project its motion upon a wall by casting a shadow of it from a small source of light some distance away at its own height from the floor. The motion of the shadow will be that of the projection of P and will be like the motion of M; the circular path of the ball is the circle of reference.

Velocity and acceleration in simple periodic motion. To find the actual values of the velocity and acceleration of the point M

(and hence the law of this type of motion) we note that the displacement of M is always the horizontal part (assuming AB to be horizontal) of the displacement of P; or, in symbols, if x is the distance CM, r is the radius of the circle, and θ is the angle PCM,

$$x = r \cos \theta.$$

If ω ("omega") is the angular velocity of P expressed in radians per second, the linear velocity v is equal to $r\omega$ (p. 93), and the acceleration of P has been found to be (p. 55) v^2/r and to be directed always toward the center of the circle. By proportion we may find the horizontal components of these quantities, which are also the velocity and acceleration of the point M.[1] The results are that the velocity is $- r \times \omega \times \sin \theta$, and the acceleration $- \omega^2 x$. The minus sign in the velocity signifies that it is directed toward the left in those parts of the motion where $\sin \theta$ is positive, and toward the right otherwise. The minus sign in the acceleration implies that the acceleration is always toward the left when the displacement is toward the right, and conversely. These results are just what we should expect. If one watches a pendulum, one notes that it slows down at the end of each swing, where for an instant the velocity becomes zero; when the point P is passing through A (Fig. 16–3) $\theta = 0$, and necessarily $v = 0$. At such a point the acceleration of P, which is along the radius (p. 56) is equal to that of M; so that when $v = 0$, it must follow that the acceleration has its greatest value. It seems odd at first for a body to possess acceleration at a moment when it has no velocity, but at that moment its velocity is rapidly changing from positive to negative, or the reverse, and the rate of change of velocity may be large. The angular velocity, ω, is defined as the angle (in circular measure) described per second. If P goes once around (2π radians) in a time T (the period), $\omega = 2\pi/T$, and it

[1] The non-mathematical reader will not insist on this process being carried through in the text. Anyone with a slight knowledge of the differential calculus will see that the velocity is

$$\frac{dx}{dt} = - r \sin \theta \times \frac{d\theta}{dt} = - r\omega \sin \theta$$

and the acceleration

$$\frac{d^2x}{dt^2} = - r\omega^2 \cos \theta = - \omega^2 x.$$

is convenient now to replace ω altogether by this new value and use for the velocity and acceleration the values

$$v = -\frac{2\pi r}{T}\sin\theta = \frac{2\pi}{T}\sqrt{r^2 - x^2}, \quad \text{and} \quad a = -\left(\frac{2\pi}{T}\right)^2 x.$$

In these new expressions no mention is made of the point P, or of its circle of reference, an idea which we need no longer retain, since the period is the same whether we think of P or of M. The acceleration expression takes a useful form when rearranged into $T = 2\pi\sqrt{-x/a}$. We need not fear that the minus sign under the radical will make it an imaginary quantity; x and a being always of opposite sign, the square root must always be a real quantity. With this understanding, we shall in future omit the minus sign.

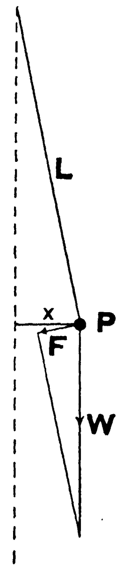


Fig. 16–4

Force in simple periodic motion. The last paragraph led to the result that the acceleration is proportional to the displacement for the type of motion there considered. Newton's second law of motion states that the force is always proportional to the acceleration; hence the force also must be proportional to the displacement in the case of a body moving as the point M does, in Fig. 16–3. In other words the motion derived by the projection of uniform circular motion follows the same law as that of elastic vibrations; both are simple periodic motions.

The force in such motion may be written

$$F = ma = m\left(\frac{2\pi}{T}\right)^2 x$$

when F is measured in absolute units; or in any units

$$\frac{F}{W} = \frac{a}{g} = \frac{\left(\frac{2\pi}{T}\right)^2 x}{g}.$$

These formulæ may be used in the solution of interesting mechanical problems.

The period of a pendulum. As an example of the use of the formulæ just obtained let us calculate the period of a pendulum. If Fig. 16–4 represents a small body P swinging on a cord of length L, the weight W is supposed not to be able to stretch the

cord; all it can do is to move P in a direction perpendicular to the cord. The component of the force tending to do this is F in the figure, acting leftward and slightly down. If the motion of P is small, its displacement x will form a triangle with L, as drawn, which will be similar to the triangle containing F and W. Hence

$$\frac{x}{L} = \frac{F}{W} = \frac{a}{g},$$

by the ratio form of Newton's law; or

$$\frac{x}{a} = \frac{L}{g}.$$

Since the period is given by the general formula

$$T = 2\pi\sqrt{\frac{x}{a}},$$

we have in this case

$$T = 2\pi\sqrt{\frac{L}{g}}.$$

This is known as the pendulum formula, showing that the period of the pendulum varies as the square root of its length. It may also be written

$$g = 4\pi^2\frac{L}{T^2},$$

in which form it shows that by measuring L and T, (both simple operations to carry out accurately) a good value for the acceleration of gravity may be obtained. This simple way of finding the constant g is the one almost always used, though for the most accurate work a less simple form of pendulum is preferred.

Newton established the fact that the period of a pendulum does not depend on the amount or kind of material in the bob. A larger inertia in the bob would be expected to slow down its motion unless the moving force (a component of its weight) increased in the same proportion. The fact that no change occurs means that ***mass and weight are strictly proportional to each other,*** and this experiment with the pendulum is one of the most accurate ways of establishing this very fundamental fact. Bessel [1] in 1830 made

[1] Friedrich W. Bessel (1784–1846), a German astronomer, now best known to physicists through his invention of the mathematical functions named after him.

still more accurate tests with the same results, using hollow bobs which could be filled with different materials.

Pendulum clocks. The pendulum clock was devised by Huygens (1658) and is based on Galileo's discovery that the period of a pendulum does not vary with its amplitude, at least if the amplitude is small. It has had an elaborate development.[1] Of the many devices involved in it only one will be mentioned here. The compensation of the pendulum for temperature changes is an important refinement. The rod may be made of invar, whose expansion may be made negligible, or, if of iron, it is furnished with a bob in the form of an iron bottle containing mercury. The mercury expands with rise of temperature. The resulting rise in its center of gravity may be made equal to the lowering of the whole bob by the expansion of the rod. Hence the center of gravity of the bob remains at a constant distance from the point of support, so that the period does not change with the temperature.

Quite recently the "Shortt" clocks have furnished us with timepieces which are more uniform in their behavior than the earth itself. By radio comparisons among clocks in different parts of the world, and by comparisons between these and clocks run by piezo-electric oscillations, whose rate is independent of gravity (p. 320), we may hope to learn much about the small irregularities of the earth's motion. The improvement in pendulum clocks has come from making the pendulum more nearly free throughout its swing than ever before.

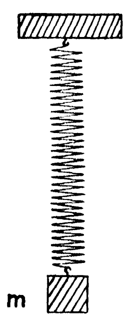


FIG. 16–5

Oscillating spring. Another interesting application of the laws of vibratory motion is furnished by a light stiff spring holding a mass m (Fig. 16–5). If the mass is pulled down from its position of rest and then let go, it will oscillate with a period which can easily be found. The general equation for the period is

$$T = 2\pi\sqrt{\frac{x}{a}}.$$

Here we must find an appropriate value for x/a. The stretching of the spring is governed by Hooke's law as in the case of all elastic bodies. If F is the

---

[1] The reader who likes ingeniously contrived mechanisms will be interested to read such an account of clocks and their development as is given in the Dictionary of Applied Physics, Vol. III, page 202, or in the Encyclopedia Britannica.

force producing the change x, we may write Hooke's law as $F = kx$, where k is some constant. k may be called the *stiffness* of the spring; since $k = F/x$, it represents the force required to produce a stretch of 1 cm., which will be larger the stiffer the spring. Now k may be measured by putting known weights on the spring and observing the resulting stretch. If $F = kx$, F is the elastic force exerted momentarily by the spring at the moment when it is stretched out a distance x, and it is also the force moving the mass m according to Newton's law, $F = ma$. Hence we may put

$$ma = kx, \quad \text{or} \quad \frac{x}{a} = \frac{m}{k}.$$

Hence the period of the spring is

$$T = 2\pi\sqrt{\frac{m}{k}},$$

which can readily be calculated. In practice it is to be remembered that if we measure F in absolute units in one case, we must do the same in the other also; that is, as deduced here, k must be the force *in absolute units* required to stretch the spring by unit length. Also, if in reality the spring is not light, but has a mass of its own, this must be take into account.

Complex vibrations. In reality simple periodic motion is not very common. It is usual to find that vibratory motions are complex. An example of such a motion can easily be produced by hanging two or more pendulums in tandem, as in Fig. 16–6. Each pendulum bob is chosen to have a weight small in comparison with that of the one next above it. If the lowest is set swinging, then the next, and so on up, the resulting motion of the lowest one will be a mixture, the sum of the motions of each, and will appear to be quite hopelessly complicated, if the observer does not know how simply it has been produced. It usually happens in nature that bodies can vibrate in a number of different ways at once, and it is also true, as in the case just given, that the motion can be analyzed into comparatively simple elements. This analysis is sometimes very useful, especially in applications in the realms of sound and electricity.

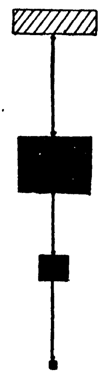


FIG. 16–6

Harmonics and overtones. A useful example of the many possible modes of vibration possessed by a body is furnished by a long elastic spring or rope, one end of which is fixed to a wall while the other is held in the hand. It is possible by moving the end of the spring up and down at a certain rate, to establish a motion of the spring as a whole, as in Fig. 16–7*a*. Twice as fast a rate gives the condition *b* and three and four times as fast give *c* and *d* respectively.

For the present, until we examine into the cause of this subdivision (p. 256), we shall take these as experimental facts. The spring is capable of vibrating as a whole, or in a subdivided state, giving a ***fundamental*** vibration, and a series of ***harmonics*** or ***overtones.*** The word "overtones" we shall use as a general term for the higher vibrations given by a body, and "harmonics" will be restricted to mean a set of vibrations whose rates, or frequencies, are in the ratios of the natural numbers, 1 : 2 : 3 : 4, etc., as in Fig. 16–7. This set of simply related vibrations is known as ***the harmonic series.*** The overtones of many vibrating bodies are not harmonic (see examples, pp. 237, 268). Any linear body, such as a wire, a rod, or a long air column in a pipe, is capable of executing harmonic vibrations, and, if disturbed without especial care, will do so, ***executing several at once.*** Thus each particle will be in complex motion, though of a sort which may easily be analyzed into its separate components.

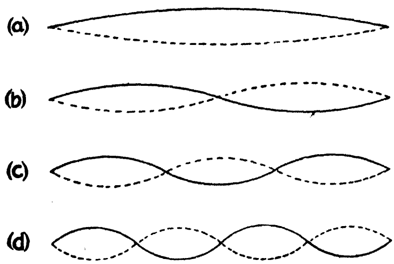


FIG. 16–7
A harmonic series of vibrations

Nodes. If a body is vibrating in such a manner that it produces one overtone only, there are places in the vibrating body which are at rest. These are called ***nodes*** if they are points, as in the case of a string. Between these are vibrating ***segments,*** or ***loops,*** in the middle of each of which lies an ***antinode.*** In the case of vibrating plates the nodes take the form of ***nodal lines,*** and the segments are vibrating areas (p. 238). Referring again to Fig. 16–7, it is seen that adjacent segments are always opposite in phase, and that the material at a node is more twisted or distorted than it is anywhere else, though it is not moved bodily; while at an antinode it moves considerably, but is not distorted.

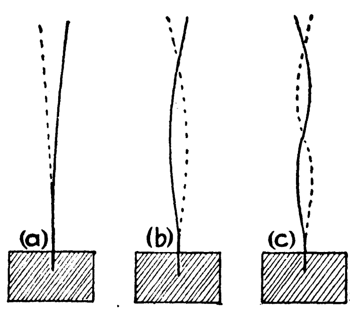


FIG. 16–8
Fundamental vibration and overtones of a rod fixed at one end

Vibrations of rods. Rods, or strips, may be supported in various ways. If fixed at one end, as in the reeds of certain organ

pipes, or of the clarinet, they may vibrate as a whole (Fig. 16–8*a*) or with one or more nodes, as in *b* and *c*. Of course the drawings grossly exaggerate the motion. The overtones in this case are not harmonically related. If the rods are supported at two suitable points, as by resting them on wedge-shaped wooden pieces, they may vibrate as shown in a much magnified way in Fig. 16–9. Such vibrations are utilized in the xylophone, and in other instruments. Their overtones are in the ratios 1 : 2.756 : 5.404 : 8.933, etc. and are thus far from harmonic.

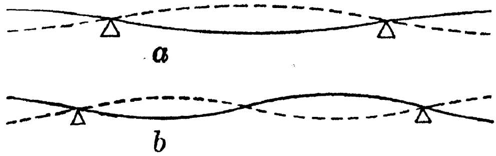


FIG. 16–9

Vibrations of plates. If a horizontal square metal plate is firmly fixed at its center, and held with certain points fixed on the rim, while it is bowed at

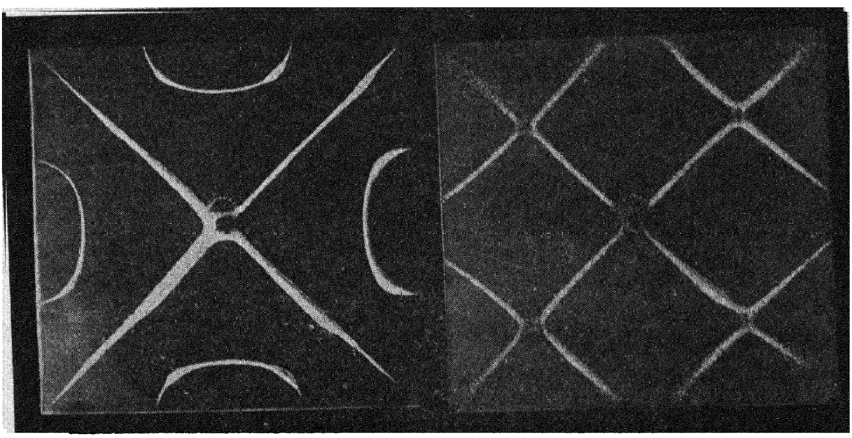


FIG. 16–10
Vibrations of plates

others, its surface may be made to break up into a number of curious patterns, and fine sand sprinkled over it will collect along the nodal lines, making these very evident. Two such patterns, known as Chladni's figures, are shown in Fig. 16–10.

The vibrations of circular plates, attached around the edge, often loaded more or less near the middle, are of interest on account of their use in telephones, phonographs, etc. Two sand figures for a telephone receiver diaphragm are shown in Fig. 16–11. These figures are interesting because they show how complicated may be the state of vibration of a body of simple form and give a vivid picture of such vibrations. The vibrations of more complex bodies (for instance, the crank shaft of an automobile engine) are often very important and exhibit some of the same general features.

Bells act somewhat as plates, and vibrate in segments. A good bell is one

in which most of its prominent vibrations sound well together. Its overtones are usually not in the harmonic series, and the fundamental tone is so low as to make a nearly inaudible hum.

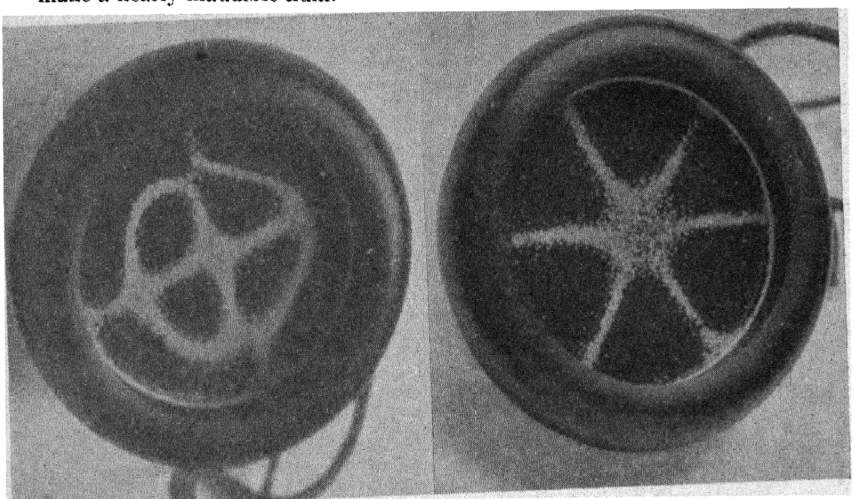


FIG. 16–11

Vibrations of the diaphragm of a telephone supplied with simple oscillating currents

PROBLEMS

1. What are the frequencies of the harmonics of a piano string whose slowest vibration is at the rate of 250 per second?

2. Find the acceleration of a pendulum whose length is 245 cm. when it is (*a*) at the center of its swing, (*b*) 5 cm. out from the center. What difference does it make in the answer to (*b*) if the pendulum is temporarily at the 5 cm. mark on its way out to a greater distance, say 10 cm., or if the 5 cm. is as far as it goes from the center?

3. A light spring supports a 100-gram weight, as in Fig. 16–5, and has such a stiffness that 10 grams additional weight would stretch it 1 cm. farther. What is the period of vibration of the spring with its 100-gram load?

4. A 140-lb. man stands on a platform supported by long heavy springs. The platform oscillates up and down with a period of 10 sec., and an amplitude of 4 ft. Find the greatest and least force he exerts on the platform, and show where each occurs.

5. A 150-lb. man sits on a wagon which is going over a road with a wavy surface, so that he moves up and down with simple periodic motion, and an amplitude of 1 ft. What is the fastest rate of oscillation which will enable him to stay in contact with his seat continuously?

6. A painter's staging is hung from the roof of a building by means of two equal ropes. If the length of the ropes (measured to the center of gravity of

the suspended weight) is 16 ft., what is the period of the staging, acting as a pendulum? If it is swinging with an amplitude of 2 ft., what is its velocity in the middle of its swing?

7. The end of the prong of a tuning fork vibrates with an amplitude of 1 mm., and a frequency of 380 per second. What is its maximum velocity?

8. A clock has a pendulum which "beats seconds" exactly (i.e., has a period of 2 sec.) in a place where $g = 979$ cm./sec.^{2} Will it gain or lose when sent north to where $g = 981$, and how much (in seconds per hour)?

CHAPTER 17

WAVE MOTION

Wave phenomena. From childhood days we have all watched with interest the beautiful aspects of waves and ripples which appear when a water surface is disturbed. A student of physics who wishes to run down and explain every fact connected therewith finds himself face to face with some very considerable difficulties. A holiday by the shore of a lake may raise many questions. What is the reason for the peculiar and complicated pattern of ripples associated with the motion of a boat? Why do these waves travel at all? What determines their speed? Why should they choose the particular rate of vibration that they do? Why does one set always form ahead of the bow of a moving boat? How is it that two sets of waves can cross each other and still preserve their form and identity? How does the wind manage to start waves at all? Why should a film of oil on the water affect the waves? Why does a ring of waves form when a stone is dropped into still water, and why does the number of ripples increase in it as the ring grows? Why does any particular ripple on which we fix our attention get ahead of the group to which it belongs and die out, the ones behind growing in prominence to take its place, so that the group travels more slowly than the individual wave? How do waves manage to spread around the corner of a pier, or

how are they reflected by a solid wall? Why do waves usually come in parallel to a shelving shore, no matter in what direction they are traveling outside? These are a few of the questions that may occur to one under such circumstances. The explanations of these facts are not all simple; some of them are beyond the scope of this book; but a few such matters will be considered in this chapter and the next.

What travels in a wave. A wave motion involves two features. One of these is a vibration of the parts of the medium in which the wave occurs, (a water surface or a long rope, for example); the other is a new feature, the traveling of a state of disturbance through or along the medium. The state of disturbance is the only thing that travels; the particles of the medium merely oscillate back and forth, or up and down, or around in a little closed curve. The medium is distorted as the wave passes through it; a water surface is partly elevated, and partly depressed, a rope is bent, etc. The distortion involves energy, and the energy travels with the wave; and the process of traveling is in most cases accomplished without appreciable friction. The waves from a passing steamer dash with great vigor on the shore a mile away; a shout starts a sound wave audible in still air at a great distance. In such cases the waves must agitate very large amounts of material on the way,[1] which temporarily acquire the energy of the waves, but pass it on again to the adjacent parts of the medium without serious loss.

New terms. A water wave involves crests and hollows; many other types of waves do not, but it is convenient to use the most familiar kinds as examples. There are two quantities connected with waves which we did not have to consider when discussing vibrations, namely velocity and wave-length. The ***velocity*** of a wave can be defined as the velocity of its crest, if it has one; or if there is no crest, as in sound waves in air, it is the velocity of some particular condition which travels with the wave, such as a maximum compression, for example. There are cases, as noted above, where this "wave-velocity" is different from the velocity of the group of waves ("group-velocity"), but these cases will be omitted for the present. The ***wave-length,*** λ, (lambda) (see Fig. 17–1) is the distance at any instant from crest to crest, or hollow to hollow;

[1] A shout audible 100 yards away agitates over 4500 tons of air on its journey.

but, to avoid terms applicable to certain types of waves only, it is better to say that the ***wave-length is the distance from any particle in the wave to the next particle which is vibrating in the same phase*** (p. 230). It is assumed that the distance is taken in the direction perpendicular to the ***wave-front,*** which is the line (or surface) which would be traced by joining ***adjacent*** particles which are in the same phase.

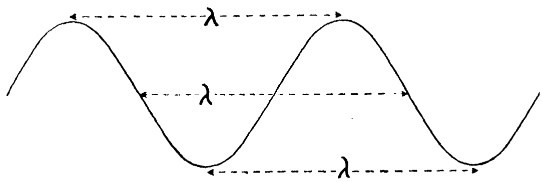


FIG. 17–1

Velocity in terms of frequency and wave-length. The velocity, frequency of vibration, and wave-length are directly related to one another. If a vibrating body, making f vibrations per second, is allowed to vibrate for one second only, it sends out f waves, each of length λ, and together they reach out a distance V, the velocity of the wave-motion. Hence $V = f\lambda$. This expression is useful, as it enables us to calculate any one of these three quantities, if the other two are known.

Water waves. What is it that determines the rate at which ordinary water waves travel? The answer to this simple question is disappointing because it is so complicated; but, on the other hand, it brings out a number of facts that are often overlooked even by those who are very familiar with the behavior of water surfaces.

As the water in a wave oscillates, the material in each crest must fall down to create the future hollow at that point. Its momentum carries it past the normal level and then when the hollow is formed, the adjacent crests tend to fall into it again. The rate at which these motions occur must depend on the speed of the waves, and a number of other factors, such as the weight of the water, the wave-length, and perhaps the depth. The theory cannot be given here. The results are that in very *deep water* $V = \sqrt{\frac{g\lambda}{2\pi}}$, where g is the acceleration of gravity and λ the wave-length; that is, the velocity varies as the square root of the wave-length. Actual values for ocean waves are 26 miles per hour for waves of 300 ft. length; 15 miles per hour for 100 ft. length, etc. The longest waves are thus seen to possess very high speeds. On account of this change of speed with wave-length, the form of the complex waves occurring in the ocean is constantly altering, the longer waves overtaking the shorter ones.

As the depth of the water diminishes, the law of speed changes, until when it is small compared with the wave-length the velocity becomes independent of the wave-length (excluding cases of very short wave-lengths) and is given by $V = \sqrt{gh}$ (for *shallow water*), h being the depth. For *very short waves* ($\lambda < 1$ cm.) surface tension becomes the controlling factor, and then

$V = \sqrt{\frac{2\pi g T}{\lambda}}$, T being the surface tension. The shortest of these waves travel extremely fast. For *rather short waves in deep water*, a combination law holds, and

$$V = \sqrt{\frac{g\lambda}{2\pi} + \frac{2\pi g T}{\lambda}},$$

which produces a minimum speed of 23 cm./sec. at a wave-length of 17 mm. This explains a curious fact, noted in 1871 by Sir William Thomson (later Lord Kelvin), to whom we owe the theory; namely, that as a small body (a stick or fish line, for example) moves through the water (at a speed greater than 23 cm./sec.), it is preceded by a short wave and followed by a longer one, each moving at the speed of the body. If the body gains speed, the length of the preceding wave becomes shorter, and that of the following one longer. This is what would be expected if there is a wave-length of minimum speed, and these waves have lengths above and below this value.

Waves in air and their speed. What we commonly mean by water waves are those occurring at the surface between air and water, producing no considerable effect at a great depth below that surface. But there are also waves which are formed and travel entirely inside a body of water; similarly the waves which carry sound to our ears are formed and travel inside the air. To see how such waves act, imagine first what would happen if a spherical soap bubble full of explosive gas were somehow set off like a bomb. There would be a sudden creation of heat with a resulting expansion of the gas, which would compress the air ahead of it in every direction, so that a thin spherical shell of disturbance would spread out from the source. At one part of this growing wave there would be highly compressed air, whose elasticity would make it expand as soon as it could, thus passing on the state of compression to the air beyond, the air particles moving along the line of motion of the wave itself. Evidently the inertia (or density) of the air must enter into this action also. A light gas will relieve itself from a state of compression more quickly than a heavy one, and this will produce a greater speed of the wave in the light gas. It can be proved that the velocity of a wave in a gas is given by $V = \sqrt{e/d}$, where e is the coefficient of elasticity of the gas and d its density. When the proper value for e is chosen, this formula leads to correct results. The coefficient e which is to be used here is the one corresponding to an adiabatic compression (no gain or loss of heat), since the compressions and expansions occur so rapidly that heat has no chance to spread. Its value (p. 219) is γp, where γ for air is 1.40, and p is the pressure in dynes per square centimeter.

Such a wave would not be periodic, but if we imagine our spherical soap bubble to expand and shrink at a regular rate instead of exploding, it will send alternate compressions and expansions out into the air, which will travel with the same speed as above. If the alternations are rapid enough, these waves will produce sound. Such waves, if we could see the particles carrying them, would present an aspect something like Fig. 17–2, which is supposed to imitate an instantaneous photograph. Such a picture we have as yet no means of obtaining directly.

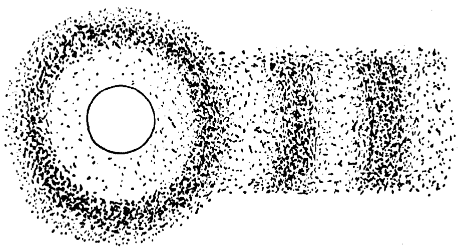


Fig. 17–2

Photography of spark waves. The nearest approach to the photograph just mentioned has been attained in the beautiful method, devised by Foley[1] of photographing sound disturbances

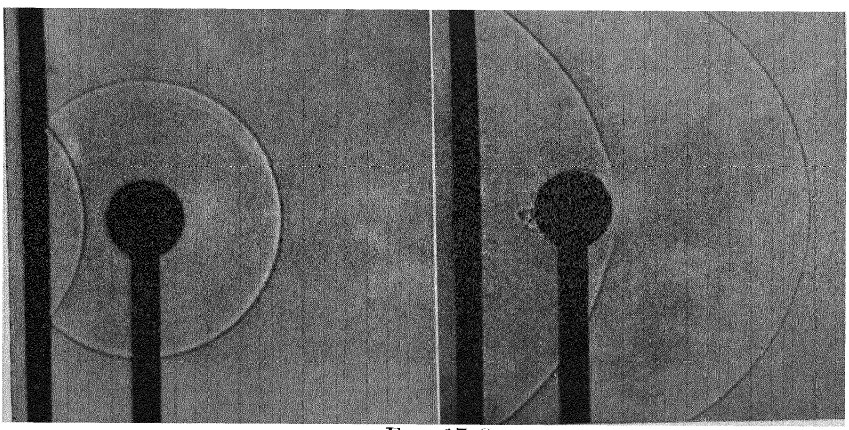


Fig. 17–3

Two stages of a reflected sound disturbance, photographed by A. L. Foley

spreading from loud electric sparks. The explosive effect they create travels outward in the air as a spherical shell, and consists of a compressed mass of air in front and a rarefied region behind. Another spark is arranged to occur on one side at an adjustable time interval after the sound wave has started. The light from this casts a shadow of the wave on a photographic plate. The "light

[1] Professor A. L. Foley of the University of Indiana.

spark" is so sudden a phenomenon that the sound wave does not have time to move perceptibly while the light lasts, and thus appears stationary on the photograph. Figure 17–3 shows such a wave being reflected from a hard plane surface. Foley has skilfully applied the method to the measurement of the speed of sound, and also to other problems.

Measurement of the speed of sound waves. The speed of sound waves in air may be obtained by firing a gun at a measured distance and timing the interval between the flash and the report. This method is direct, and simple to understand. The flash arrives practically instantaneously. The distance and the time are easily measurable with fair accuracy. For work of the highest precision, however, the method is less satisfactory, as the condition of the air through which the sound passes cannot be accurately known, and even the exact path followed by the sound (and therefore the distance) may be in doubt. A better way is to determine V from f and λ and use the equation $V = f\lambda$. This is a laboratory method, and is usually done with air in a tube or pipe, but the tube itself has a slight effect on the velocity, so that this method, though better, is not perfect.

For use in problems we may take the speed of sound in air in round numbers as 340 meters, or 1100 feet a second. The first is an accurate value at about 15° C., and the second near 7° C. At 0° C. the best value is 331.46 meters per second, or 1087.5 feet per second. It varies by 2 feet or 60 cm. for each degree centigrade, increasing with rise of temperature.

Consequences of the small velocity of sound. Curious effects are sometimes observed which are due to the fact that sound waves move on the whole rather slowly. For instance, a band may be supposed to keep all the marchers in a long parade in step; but an observer who has a good view of a long section of the parade can readily see that those who come some distance behind the band lag behind the step of those nearer to it. The conductor of a large chorus and orchestra on some festival occasion keeps all the musicians together by his beat, so that they sing and play simultaneously; but it does not sound so to a listener at one side if the space over which the performers are spread is large; those at a distance seem always to be behind. In large churches the organ is often in two parts; one, an "echo organ" being placed some distance away; if the distance is great it is impossible for

the two parts to sound simultaneous to listeners in all parts of the building.

When a lightning stroke occurs, it may be a mile or so long, and part of it may happen to lie in a line pointing toward an observer; if so, the sound, which is created simultaneously in all parts of the discharge (as far as we can determine), does not appear so; the nearer parts are heard first. The resulting sound may resemble the tearing of sheets, or, if the spark is a long one, a continuous roll of thunder is produced from a single instantaneous shock. The jaggedness of the spark accounts for the usual lack of uniformity in the loudness of the thunder. A lightning stroke a mile long, taking place more or less in the line of the observer, would produce a rolling sound lasting nearly five seconds.

Bullets and high-speed shells can be given a speed much greater than that of sound. If such a shell passes over an observer, the first sound he hears is an apparent explosion over his head, which is really a concentration of the sound made by the shell tearing through the air, from parts of the path approximately equidistant from the observer. After this he hears two sounds at once; one being that of the shell as it proceeds beyond him, the other that of the shell advancing on its way toward him. Lastly he may hear the shell start on its flight, the initial sound wave being so much behind in the race.

Sound speed not dependent on frequency or pressure. The frequency of the vibration has no appreciable effect on the speed of sound waves. This is plainly shown if one listens to a distant band; the music is heard in its proper form. If the high-frequency notes traveled much faster than the others, the music would be hopelessly distorted. Remembering that only the elasticity and density of the air were mentioned in the formula for the speed of sound, we need not be surprised that the frequency has no effect.

Through the ordinary range the pressure, likewise, has no effect on the speed of sound. Experiment and theory agree on this point, though offhand one might not expect it. The velocity formula contains the ratio of the pressure and the density; since these always change together, if the temperature does not vary, their ratio is constant. The kinetic theory also leads to the same conclusion. According to it sound waves are carried by that part of the natural kinetic agitation of the particles which is in the direction of motion of the sound waves. Since the velocity of the gas particles de-

pends on the temperature, but not on their number, the speed of sound ought to be independent of the density and thus of the pressure also.

At very high or very low pressures these conclusions no longer hold true.

Sound ranging. Exact knowledge of the value of the velocity of sound is sometimes a matter of vital importance. For instance, an interesting method, known as "sound ranging," was developed during the World War for finding the location of the enemy's guns by sound. The sound which comes from the muzzle of a big gun when the shell leaves it is known as the muzzle wave. This spreads over the ground in a circular form with center S (Fig. 17–4), and at one instant has the form AFG. At this instant it reaches an observing station A where is placed a listening microphone (p. 340), which is one of a series, of which B and C are others. A, B, and C are often placed on the arc of another circle whose center is somewhere in the midst of the enemy's guns. The receiving microphones are all connected to a central station, where the time intervals between the arrival of the sound at A, B, and C are exactly recorded by an ingenious photographic device, to a thousandth of a second. Knowing the velocity of sound and these time intervals, it is possible to draw circles, as in the figure, with centers B and C, and radii proportional to the distances traveled by the sound in these intervals. Hence we can be sure that the sound wave when it reached A must also have been tangent to these circles at the same instant, and it therefore had the form AFG. Thus the center S can be found.

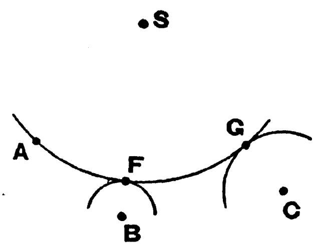


FIG. 17–4

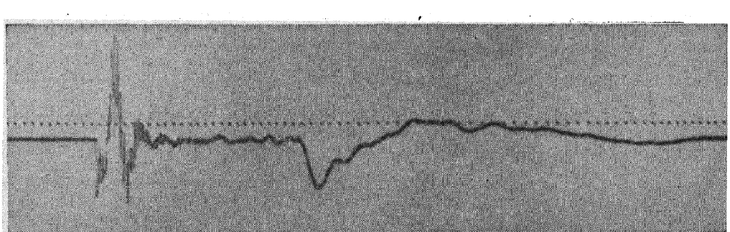


FIG. 17–5

Phonodeik record of the discharge of a 12-in. gun taken by Professor D. C. Miller

Quick and accurate methods of calculation were developed, whereby the position of the gun could be found within a time of two minutes, and with a precision as great as that obtainable by any other methods; this could be done equally well in foggy weather or at night. Precautions had to be taken,

however, to compensate for the effects of wind and of temperature variations in the air.

Figure 17-5 shows a record taken with the phonodeik (p. 262) of the sound produced by a shell fired from a high-speed gun. The first disturbance (on the left) comes from the shell overhead, the larger one from the muzzle wave. The dots record time intervals; they are $\frac{1}{200}$ sec. apart. Sound-ranging records are somewhat similar, but one film carries several traces, as well as the time marks, so that differences in time of arrival at different stations are easily read on them.

Variation of the speed of sound with temperature. As already indicated, the variation of the speed of sound with the temperature is rather large, and it is interesting to note the simplicity with which it can be explained on the basis of the kinetic theory of gases. According to this theory, each gas particle spends most of its time flying about alone. It is true that it collides with other particles very frequently, but the collisions last a very short part of the total time. In the propagation of waves through air, the small motions causing the waves are definitely directed, and these must be superimposed on the irregular ones already existing among the gas particles. Such motions can be passed along from one set of particles to another only at the rate that the gas particles are themselves moving in the direction of the sound wave. As the kinetic energy of the gas particles (which depends on V^2) is proportional to the absolute temperature, the velocity of the particles, and hence also of the sound waves, must be proportional to the square root of the absolute temperature. Hence the velocity of sound V' at 1° C. must be connected with the velocity V at 0° C. by the ratio $\frac{V'}{V} = \sqrt{\frac{274}{273}}$, which equals 1.00183, and yields precisely the observed value for this change, namely, 60 cm. per degree.

Sound phenomena out-of-doors. Variations in sound speed are observed very commonly in the open air. They are due not only to changes in temperature, but to changes in the water-vapor content of the air, and to winds. These variations sometimes produce curious effects. Water vapor is lighter than air, and thus "moist" air (i.e., air of high humidity) transmits sound faster than dry air. A following wind also speeds up the sound in it. In Fig. 17-6 a source of sound S is situated in a region where the air is cool below and warm above, as over a small lake on a summer evening. In such circumstances the speed of sound increases with the height above the surface according to some unspecified law.

The result is that the sound wave-fronts, instead of keeping the form of circles with center S (shown dotted), begin to lean over toward the listener L, and may become concave to him, in which case the path of the wave (perpendicular to the wave-front) is curved, and the sound bends over, comes down and becomes more or less concentrated upon the listener. It is well known that hearing may become extraordinarily good under such conditions. Precisely the same diagram would apply to the case of air of uniform temperature in which there was a wind blowing from S toward L with increasing speed as one ascends, which is a quite usual speed

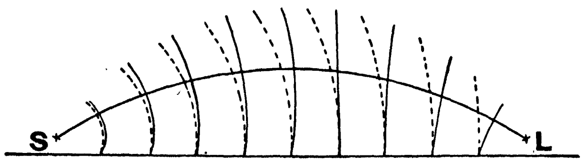


Fig. 17–6
A curved "ray" of sound

distribution on account of friction near the ground. The same effect would be produced if the amount of water vapor in the air increased with the height.

The opposite condition, when the path of the sound is deflected upward, so as to pass over the listener's head, will occur if the wind is blowing against the sound; or if there is the reverse stratification of temperature, cold above and warm below, as in the case of an open field heated by the sun's rays; or if the air below contains more moisture than that above, as it often does. These cases are similar to mirages in the case of light (p. 488).

As such phenomena are very common, it becomes a difficult matter when one hears a distant sound to say whether the sound followed a straight path from the source to the listener, or a curved one. As the curved one is the more likely, the measurement of the speed of sound in the open air becomes somewhat uncertain on this account if the utmost accuracy is desired.

It may even happen that sound in passing over land and water in succession, may be bent first up and then down, so that the noise of a battle, or even of a foghorn, may be inaudible comparatively near by, and audible again at great distances.

As an irregular distribution of moisture, temperature, and wind is a commoner arrangement than stratification, sound usually

finds a varied and inhomogeneous atmosphere through which to travel. This makes the progressing wave-front become crooked, so that different parts move off in different directions; thus the sound becomes scattered and cannot keep its wave form, and hearing at great distances becomes impossible. Such conditions are very prevalent on clear, warm days, when a very loud sound may be inaudible a mile off, though on a still, foggy night, when the air is uniform, it may be heard ten miles away from the source. Precisely the same conditions produce effects on light waves which cause the twinkling of the stars when seen near the horizon.

Longitudinal waves in fluids. Such waves as have been discussed above are called compressional or longitudinal waves. In them the individual particles move back and forth with periodic motion in a short line in the direction of motion of the wave as a whole. In the interior of a fluid (i.e. a liquid or a gas) this is the only possible sort of wave motion. This motion can occur only when the deformation associated with it is resisted by elastic forces in the material. It is evidently impossible to get any resistance in a liquid or a gas to a change of shape, or "shear" (p. 123); hence transverse disturbances (including torsional) can start no waves in such materials.

Under-water waves. Compressional waves are as common in liquids as in gases. Such waves in water are not what were called "water waves" above. Though just as frequently produced, they are less often noted by man, though fish and diving birds hear by means of them. The common water waves are surface effects. The agitation associated with them diminishes very rapidly with depth. A submarine can find calm water in a storm by submerging to a moderate depth. Compressional waves, on the other hand, penetrate the whole body of the water, and involve very small motions of the water particles, like sound-wave motions in air. They travel with relatively high speed, over four times as fast as sound waves in air, because of the great resistance which water offers to compression. Under-water waves are, strangely enough, what we all hear by, as the inner ear is filled with liquid.

Recently devices [1] have been invented for sending and receiving these waves so that they can be used for submarine signaling between

[1] For an account of these see "The Mechanical Properties of Fluids," by a group of authors (D. Van Nostrand Co., 1924), Chapter 9 by Dr. C. V. Drysdale, on Submarine Signaling.

ships, or with shore stations, or for finding the depth of the water by echoes. (One listening device is illustrated in Fig. 17–7.) It is thus possible for a ship when properly equipped to obtain practically continuous records of the depth as it proceeds, by making short, sharp under-water sounds and measuring the time it takes for them to go to the bottom and return as echoes to the ship. The velocity of the waves being known (1450 meters a second in fresh water; 1454 in average sea water, both at 0° C.) the depth is easily obtained. The speed of these waves increases by about 4 meters a

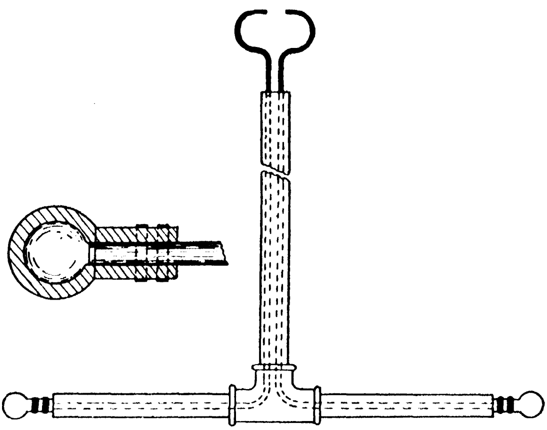


FIG. 17–7

On a framework that can be turned about a vertical axis, two hollow rubber bulbs are held under water, connected separately by pipes, each to one ear of the observer. The effect is as though the observer had ears several feet apart, under water. By turning this device the direction of a sound can be determined, just as we do by turning our heads in the air. This was called the "C tube" (after Dr. W. D. Coolidge who invented it).

second with each degree centigrade, and varies considerably with the salinity of the water.

Slow compressional waves (about 50 oscillations per second) have been used for the transmission of power along pipes full of water. They are then analogous to alternating electric currents. Interesting machines somewhat like alternating-current motors had to be invented in order to utilize this power. Further engineering developments of this sort are likely.

Transverse waves in solids. Transverse waves are of common occurrence in solids. A classical example is furnished by the sound of an approaching train which can sometimes be heard along the rails long before it is heard directly through the air. Another ex-

ample frequently encountered, and often with displeasure, is the transmission of noises along the solid framework of a house from one room to another, perhaps some distance away; or in the passage of sounds through a system of steam pipes in a building. The velocities with which such waves travel through a few common materials are given in Table XXI.

TABLE XXI

Velocity of Sound

At 0° C. in meters per second

Iron	5100	Water	1450
Granite	3950	Air	331.46
Glass	5500	Carbon dioxide	258.0
Maple wood	4100	Hydrogen	1270

The velocity in each material follows a formula of the type already given for gases, namely $V = \sqrt{e/d}$.

A large-scale example of such waves is furnished by the vibrations transmitted from earthquakes through the body of the earth, or along its surface crust. An observing station some distance from the source of the disturbance receives first longitudinal waves

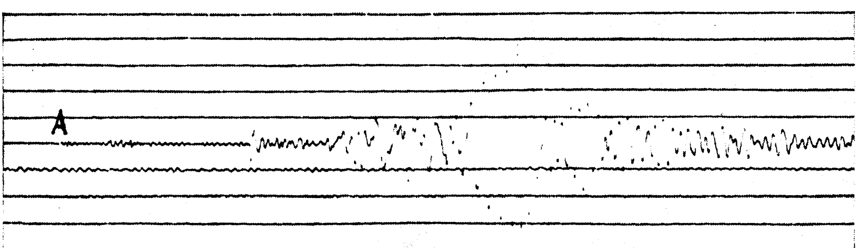


FIG. 17-8

Part of the record of the horizontal (E-W) motions of the earth at the Dominion Observatory, Ottawa, Ontario, caused by an earthquake in the Gulf of Panama on Aug. 9, 1927. The parallel lines are traces, one for each hour, broken on the original record by small gaps marking one-minute intervals. At the point marked A the disturbance began with small longitudinal vibrations transmitted through the body of the earth, followed six minutes later by larger ones due to transverse waves in the earth's crust. The most conspicuous of these have a period of 20 seconds, and are thus quite slow. They had not quite died out in two hours. The whole record is very like one showing the sound vibrations from a bell.

which have come through the body of the earth, and, after the first of these have arrived, a set of transverse earth-crust waves begins. The speeds of these two sorts of waves being known, the time between their arrival gives the distance of the earthquake

from the observing station; not, of course, its direction. The necessary computations are not difficult. Three such observations, taken at different points on the earth's surface will serve to locate the earthquake definitely. A record of the sort taken is shown in Fig. 17–8. The instrument with which this record was taken is called a seismograph. It works on the principle that a heavy mass, supported so as to be quite free to move, say in a horizontal direction, will stay at rest on account of its inertia, if the earth oscillates in this direction. If the stationary mass carries a recording pen, and the part of the instrument carrying the paper moves with the earth, a record will be made of all horizontal motions. Other instruments of the same sort are made to record vertical motions.

Reflection of waves. A wave in any material may be reflected when it reaches a boundary of that material. The amount of reflection will depend on the velocities of the waves in the two materials, none occurring if they are equal. As a mechanical example, which is readily understood, let us consider a row of balls hung on strings, as pendulums, and almost touching each other, as in Fig. 17–9, one-half of the balls being light and the other heavy. If ball A is given some motion in the direction of the arrow, it will strike B and pass on all of its motion to it, coming to rest after the collision. B in turn will pass it to C, and thus it will progress till it reaches G, the last of the larger balls. But when G hits H, the small mass of H will not be enough to stop G. H will be driven forward; G will also move forward at first, then turn and start a reflected impulse which will travel back to A. There will be reflection at the "boundary" GH if there is a difference between the two sets of balls, and this will occur whichever way the impulse is going. The impulse which G handed to H will go on to the last ball Z, which will swing out and on its return start a reflected impulse back toward A. Examples of reflection like this, at the free end of the medium, are very common. It is just as easy for reflection to occur from nothing (i.e., the end of the medium) as from a heavy rigid

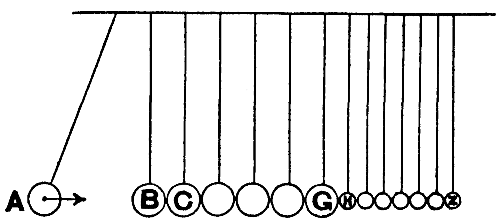


Fig. 17–9

boundary. Wave motions are reflected in much the same manner as those in this mechanical example.

In air these reflections are the cause of echoes, and of the confused reverberation which is often heard within large lecture or concert halls. Part at least of the ease of hearing over smooth water is due to the almost perfect reflection of the sound waves at the surface.

Reflection of waves from rough surfaces. Reflection of sound occurs from objects whose nature would seem not to offer the proper conditions; for instance, from the edge of a mass of tall trees, whose "surface" is so irregular that one would not be disposed to call it a surface at all. ***If the irregularities of the surface are small compared with the wave-length,*** reflection occurs as though the surface were ***perfectly smooth.*** While a group of trees does not even in this sense possess a smooth surface, it is often evident that it gives a better echo for low-pitched sounds than for high, because the wave-length is longer, and hence the surface is "smoother" in the sense of the statement above. This case is not very important; other examples of this phenomenon which are of more interest are met with in the study of light waves (p. 483).

Change of phase on reflection. A long rope, fastened at one end, can be moved by hand at the other end so as to make a wave pass along it. The disturbance sent down the rope moves at a definite speed, determined by the elasticity and density of the rope. In fact its speed is given by a formula of the familiar shape, $V = \sqrt{T/m}$, where T is the tension (the force stretching the elastic rope) and m is the mass per unit length. This disturbance is reflected at the fixed end. If the disturbance is of the form of a hump, it is found to return as a hollow, ***the phase being reversed on reflection from the fixed end*** (Fig. 17–10). If the end is free, (which can be managed in practice by tying the rope to the wall by a length of several feet of light string, so that the end of the rope can move freely up and down) a hump sent along the rope returns as a hump without

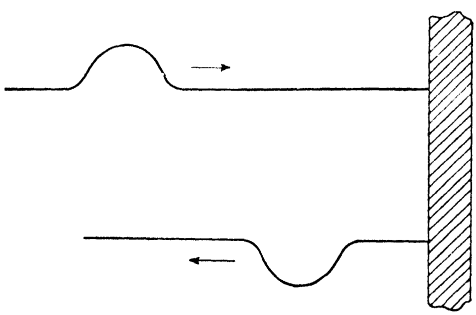


Fig. 17–10

change of phase. This fact has applications in several parts of physics, and can be proved mathematically.

Superposition of waves. If, instead of a single hump, one sends along a rope a continuous series of alternating humps and hollows, a regular wave keeps on traveling to the fixed end, and being reflected there. It is important and curious that the forward-moving waves do not in any way obstruct the returning ones. This is a special case of a very general statement that waves involving no violent disturbances can be sent through the same medium in any direction and in any numbers without mutual disturbance of form. It is easy to drop two stones on a smooth surface of water and observe the growing rings of waves interlacing with each other,

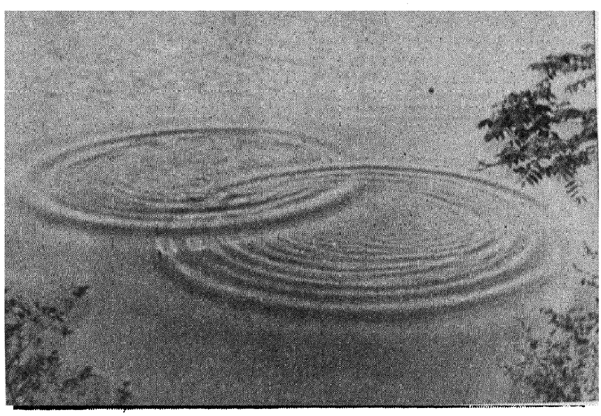


FIG. 17–11
Two stones were thrown into the water and formed these rings. The irregular rings near the centers were formed by splashes.

but retaining their identity and independence of action (Fig. 17–11). Everyone knows that the air in a room may carry more than one conversation at a time without mutual confusion, and this is so with other sorts of waves also.

Formation of standing waves. In the case of the rope just considered the forward and returning waves are carried by the same particles, and any one particle must at one instant suffer two displacements simultaneously. This means, of course, that its actual displacement is the sum of the two. If the two are equal and opposite for the particle, it is not displaced at all; if they are equal and concurrent, the displacement is doubled. If the oscillation of the hand is managed at a proper rate, the return wave meets the for-

ward one in the same phase at the middle of the rope, and at all other points the two waves meet in different phases, opposite at the ends. The result is that the rope vibrates with its fundamental vibration (Fig. 16–7) with nodes at the ends and a loop in the middle. If a rate twice as fast is chosen, it is evident that the condition just described will occur twice along the rope; for there will be two waves advancing and two returning in the distance formerly occupied by one, so that the first harmonic vibration will occur with a node in the middle, as well as at each end. Hence it is that in the harmonic series (p. 236) the frequencies are in the ratios of the natural numbers. Such waves, which no longer appear to travel at all are known as ***standing*** (or stationary) ***waves,*** and they occur whenever a wave is advancing and returning by reflection through the same medium. They are common in musical instruments.

Standing waves in air columns. In narrow air columns the same sort of thing happens. Here a disturbance entering one end of the tube will be reflected from the other and return, forming a standing wave with nodes and antinodes within the tube. If the far end is closed there must be a node at that point, for no motion can occur against a rigid wall. If the far end is open there must be an antinode there for the reason that an antinode is the place in a standing wave where large motion but no distortion occurs (p. 236). The type of distortion in air waves is compression. At an open end free motion is possible, and the openness assures the constancy of pressure and hence the absence of compression. In a standing wave so chosen that the wave-length is just the length of the vibrating body (as in Fig. 16–7*b*, p. 236), it is easy to see that the wave-length is twice the distance from one node to the next, or from one antinode to the next. Figure 17–12*a* shows the lowest and simplest type of vibration for a pipe closed at one end, producing a wave whose length is four times that of the pipe. If this pipe breaks into an overtone, it must do so by subdividing into an odd number of parts,

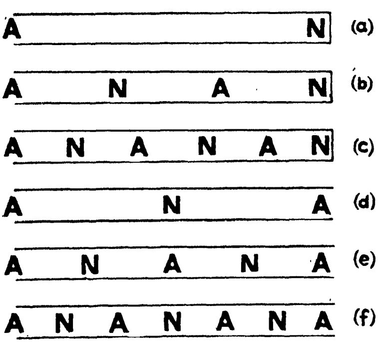


FIG. 17–12
The vibrations of air columns

since its ends are unlike. Representing an odd number by $2n - 1$, where n is any integer, the wave-length is given by $\lambda = 4L/(2n - 1)$, L being the length of the pipe. This equation gives the wave-length of each of the vibrations that can be formed in this pipe. The figure shows two more of these.[1]

If the pipe is open at both ends, there must be antinodes at the ends, and the simplest mode of division (fundamental vibration, Fig. 17–12*d*) gives a node in the middle. The overtones are found as shown by Figs. 17–12*e* and 17–12*f*. In this case $\lambda = 2L/n$ for all possible wave-lengths.

Kundt's tube. An interesting example of standing waves in air columns is furnished by the Kundt's tube experiment. It shows in a striking manner the form of the waves as they exist inside a glass tube. Usually a rod (Fig. 17–13) clamped firmly in

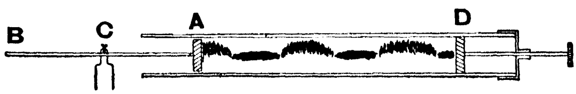


FIG. 17–13
Kundt's tube

the middle, at C, has a cork, A, on the end which fits loosely in a long glass tube, so that when the free end B of the rod is stroked with a resined cloth, it will vibrate longitudinally, the ends A and B being always opposite in their motions, and C remaining at rest. The motion of A excites the air in the tube AD, and if a line of cork dust is placed in the tube and the tightly fitting piston P is properly adjusted, a vigorous standing wave arises in AD when BC is stroked. The figures formed by the dust are more distinct if the dust is first formed into a narrow line by tapping, and then the tube is rotated a little around its own axis until the dust is almost ready to fall down. Then the dust at the antinodes falls, and the wave figure is easily seen. Figure 17–13 shows the appearance of the dust figures as seen from above.

A somewhat easier method of exciting these vibrations is to fit the mouthpiece of an ordinary wooden organ pipe to the glass tube. The piston D may be omitted, and then the natural har-

[1] The nodes and antinodes here given refer to *displacements*. If we were interested in the distribution of *pressure* along the tube we might represent it by a series of "pressure nodes" and "pressure antinodes." A pressure node is a place where the pressure does not change; i.e., it is a displacement antinode. Thus the lettering in such diagrams is the reverse of that given here.

monic series of the tube can be obtained, each with its appropriate pattern of dust figures. In this form the experiment gives a vivid picture of what is really going on inside such instruments as the flute or the clarinet while they are being played.

Measurement of the speed of sound in metals. The Kundt's tube experiment furnishes an experimental method of measuring the velocity of sound in metals. The distance from node to node in AD being measured, one can find the wave-length in air, λ_a, which is double this amount. The vibration in the rod itself is really part of a standing wave in the metal, with a node at C and antinodes at B and A. Hence the wave-length of sound in the metal, λ_m, is twice the length BA. If V_a is the speed in air, and V_m the speed in the metal, $V_m = f\lambda_m$ and $V_a = f\lambda_a$. The frequency being the same in both cases, $V_m/V_a = \lambda_m/\lambda_a$. Hence V_m can be found.

Overtones of horns. Another vibrating air-body of practical importance is found in the horns, used as amplifiers in the megaphone and the phonograph, or in some other cases as receivers. This form of air chamber can vibrate in a complicated series of overtones, and will in practical use produce an excessive amplification of those musical notes whose frequencies happen to coincide with its own overtones, with resulting distortion in the sound. More power may be sent out at these frequencies on this account, and if the horn is a long one, the overtones lie close together and the effects of adjacent overtones overlap somewhat, so that practically the whole range of the musical scale becomes amplified, and the horn is very useful. Thus in the commercial reproduction of sounds (in theaters, etc.) horns over 12 ft. in length are often used. The use of horns as receivers is disappointing, even when they are large; a considerable part of the sound energy is so reflected inside the horn that it emerges from the large end by which it entered. Nevertheless, large horns have been successfully applied to the detection of distant airplanes, etc.

Interference. Standing waves offer an example of ***interference,*** at least at the nodes. It is possible to combine two wave-motions in such a way that they destroy each other, the necessary condition being that they shall be opposite in phase. The energy is not destroyed, but is shifted to some near-by place. In standing waves interference occurs permanently and completely at the nodes, and partially near them, the energy being shifted to the antinodes. An interesting example is furnished by a room with hard reflecting walls. If a sound is started in front of a wall, a set of standing waves is formed between the source and the wall, which can be observed if the wave is simple and sustained. Usu-

ally, however, each wall is capable of producing a standing-wave pattern of its own, as are also the ceiling and floor; as they all share the same space, these waves combine into a three-dimensional pattern in the room, which is a compromise among them all. If the observer moves his head in such a room he hears variations in loudness, often finding an increase in loudness by going farther away from the source, contrary to experience out-of-doors. The sound appears loudest when the ear is at a node where there is no motion but the greatest pressure change, for it is the pressure change that drives sound waves up into the tube of the ear.

Beats. In another experiment fluctuating interference can be produced (Fig. 17–14). If, for example, we send out into the air two sound waves whose frequencies are 250 and 251 per second, they may agree in phase at one instant and produce a loud sound at the observer's ear; but half a second later they will be opposite in phase, and destructive interference will occur there. At the end of a whole second each will have executed a whole number of vibrations, and they will again be in phase with each other. In this example the sound will fall and rise once in a second; if the frequencies had been 250 and 252, there would have been in the same time two of these variations in loudness, which are called ***beats.*** In general, ***the difference in frequency between the tones is equal to the number of beats*** heard per second. This fact makes it possible by counting beats to find the exact pitch of a tone which is near one we know already, provided that we know which of the two is the higher. We can also tune two notes to the same pitch with extraordinary precision by listening to beats. One beat in ten seconds is easy to observe; two sounds which produce this effect differ by only one-tenth of a vibration per second. If their frequency is about 1000 per second, this amount of disagreement corresponds to the extraordinary accuracy of one-hundredth of one per cent, (1 in 10,000).

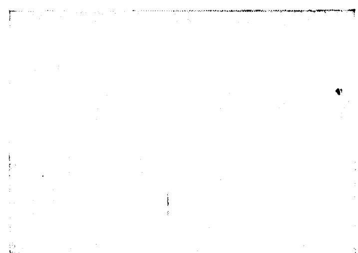


FIG. 17–14
Rapid beats as shown by the phonodeik

The observer listening to beats in an ordinary room is likely

to notice that the sound does not come down to a dead silence in each beat. This may be because the two generating sounds are unequally loud, or because the ear receives not only the direct sound but reflected sounds also, some of which have traveled a longer path, and therefore differ in phase; hence complete destruction of all the sound cannot occur at the same instant. Nor do all the listeners in a large room hear the quiet moments at the same times.

Quality of sounds. Even such persons as are not gifted with very "musical" ears may detect the differences in tone ***quality*** among such musical instruments as the flute, clarinet, and violin. In all these the vibrations consist of a fundamental tone and a series of overtones. The relative proportions of each vary from instrument to instrument. The quality of the sound is due to the relative prominence of the fundamental and the various overtones which are present. A pure tone consisting of a simple periodic vibration is unusual; it is given by a tuning fork, gently struck by a soft hammer, or by a flute or a French horn, softly blown, and by certain weak organ pipes. Strident or harsh tones are those whose high overtones are prominent. A sound which is not periodic at all is called a ***noise,*** while a periodic vibration gives a ***tone,*** or a musical note; these grade into each other with no sharp boundary.

In actual practice we recognize musical instruments not only by the quality of their tones but by the manner in which these start and stop, or change in volume. Running a phonographic record of piano music backward (which can sometimes be done by hand) gives an impression of a quite different instrument, though the quality, as defined above, must be the same. Musicians habitually argue over the tones that can be obtained from a piano; the artist's "touch," they say, can vary the quality. An inspection of the striking mechanism of a piano shows that the string may be struck a hard blow or a gentle one, but no other variations are possible. The quality varies somewhat with the strength of the blow, but the great differences involved in the artist's touch come from other features of the performance.

Analysis of sound curves. If the air is made to vibrate in five simple periodic motions at once, these motions may be exhibited separately as in (*a*), (*b*), (*c*), (*d*), and (*e*), Fig. 17–15 and these may be combined, as in (*f*), by adding the displacements

(ordinates), whose sum will then represent the actual motion of the air particles under these circumstances. This summation was performed mechanically in this case, but can be done otherwise. The actual form of the air vibrations can be observed by a method about to be described. If a curve like (*f*) is obtained, we see at once from the figure the correct analysis of this curve into its constituents. ***Whatever form of curve is found, it can be analyzed into a series of harmonic curves.*** This is known as a Fourier[1] analysis, and there are ingenious ways in which it may be performed mechanically.[2] Thus the peculiar quality of a musical note can be examined and a complex vibration analyzed into its components. As an example, a particular note (C, 270) of a clarinet has been found to possess a relative strength among the component harmonics indicated by the following order: the strongest was the fundamental tone 1, then 3, 10, 9, 8, 11, 5, 12, etc. (2 and 4 being absent). Such analyses have practical uses in the making and improvement of musical instruments and radio apparatus, in increasing the efficiency of foghorns, and also in the study of machines for generating alternating electrical currents.

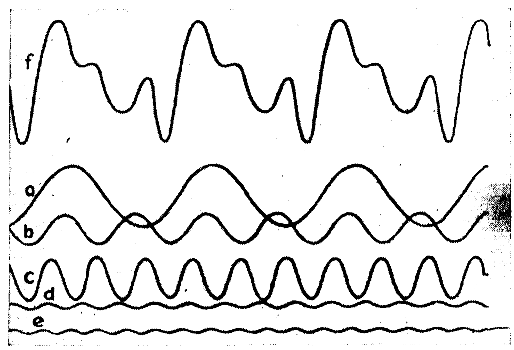


FIG. 17–15

The phonodeik. Curves showing the actual vibrations of the air particles can be obtained by a type of instrument which we owe largely to Miller.[3]

---

[1] J. B. J. Fourier (1768–1830), French mathematician and politician, who in spite of humble birth and poverty became a professor in Paris and a member of the French Academy. His mathematical theory on the flow of heat is his most famous work.

[2] The reader is referred to D. C. Miller, "The Science of Musical Sounds," 1916 (Macmillan), where he will find the whole subject on the study of musical quality by physical methods authoritatively and attractively presented; also H. Fletcher, "Speech and Hearing," 1929 (Van Nostrand), which gives an interesting account of the results obtained in the Bell Telephone Laboratories in this field.

[3] The form here described contains improvements due to Professor S. H. Anderson of the University of Washington.

The general idea is that the mass of the vibrating air is sufficient to move a light diaphragm with a multiplying attachment, which renders these motions visible, and translates them into a wave form.[1] Figure 17–17 shows the vital part of the actual apparatus, and Fig. 17–16 a simplified sketch illustrating its action. A horn H gathers up a considerable volume of sound which forces a thin glass or mica diaphragm, D, to vibrate back and forth. Attached to this is a very light steel fork, which rests on a thin straight steel pin, P, on the middle of which a bit of light mirror is fastened. A small horse-shoe magnet, not shown in Fig. 17–16 pulls the fork and the pin down (i.e., perpendicular to the paper in the figure) upon its pole faces, on which the pin is free to roll, as the fork and diaphragm vibrate. A round pencil on a table can be rolled in a like manner by resting one's hand flat upon it, and moving the hand slightly back and forth along the table. This rolling of the pin rocks the mirror through a small angle, and a beam of light directed toward it will be reflected and can be received as a spot on the screen, which moves up and down as the diaphragm oscillates. The light on its way to the screen is actually reflected from a rotating mirror, R, usually four-sided, and is thus made to trace out a horizontal line on the screen when there is no sound in the room, or a combination of this line with the vertical vibration when a sound is made. The wave form thus seen on the screen is a more or less faithful copy of the actual wave form of the air vibrations. If the diaphragm, or the air in the horn, prefers to vibrate at certain rates, as is usual, these frequencies will be somewhat overemphasized in the record, but such defects may be corrected. The records, especially when taken directly on a moving photographic film (rotating mirror then omitted) are very useful in the study of sound quality, and in other ways. Figure 17–18 was made in this way. Small changes in quality show much more delicately than they do to the ear. The instrument may be made to magnify over 40,000 times; the geometry of the arrangement shows that the magnifying power is the ratio D/r, where D is the distance from the phonodeik to the screen, and r is the radius of the little rocking pin. From the amount of motion

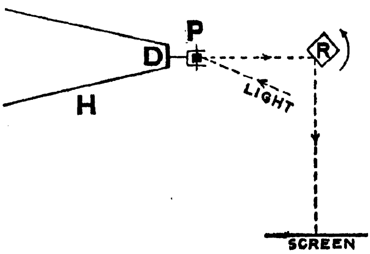


FIG. 17–16

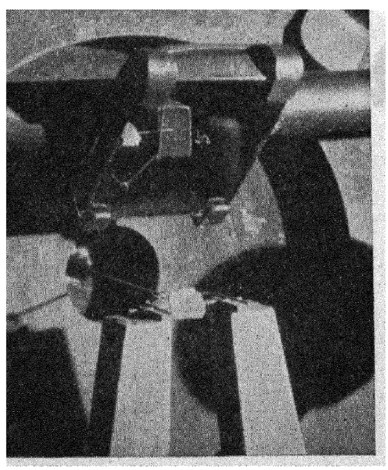


FIG. 17–17

[1] There are other ways of doing it, using electrical apparatus, microphones, amplifiers, and oscillographs.

of the spot of light that of the diaphragm can be found, and hence that of the air itself can be estimated. Thus we find that when the air particles move only 0.1 mm. a loud sound is produced.

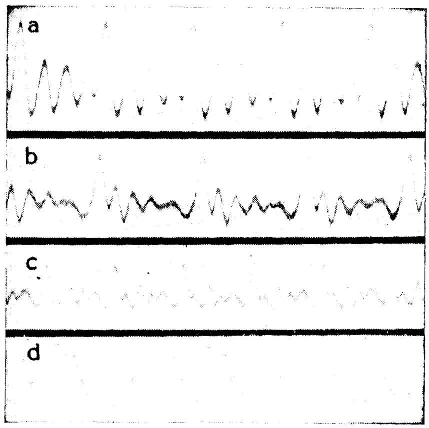


FIG. 17–18

The upper curve is a phonodeik record of the vowel *o* (as in "go"), intoned at a pitch of about 120 per second. The other three are all of the vowel *a* (as in "part") as uttered at different pitches and by different people.

PROBLEMS

1. If the speed of sound is 340 m./sec., what is the wave-length of the tone of a man's voice whose frequency is 170 per second, and of the shortest wave he can hear, 20,000 per second?

2. Find all the frequencies that can be emitted by an organ pipe 2 ft. long, closed at one end.

3. An open organ pipe is 10 cm. long. Find the frequency of the fundamental tone that it emits, and the frequencies of its overtones.

4. Prove the expression for the magnifying power of the phonodeik, given on p. 262.

5. Find the speed of a ship in mid-ocean that can just keep up with a wave 200 ft. long.

6. When a loud high-pitched note, as from a small organ pipe, is sounded in a room with bare walls, one may notice great changes in loudness by moving quietly about in the room. Explain.

7. What sort of waves can be transmitted by a fluid? Explain why not all sorts.

8. A tuning fork made of invar has a frequency of 435 per second, which is practically independent of the temperature. Explain exactly how the tem-

perature of a church could be determined by counting the number of beats per second between this fork and a pipe in the organ which gives 435 vibrations per second at 15° C. (Assume the length of the pipe itself to remain constant.)

9. In a Kundt's tube, the dust heaps in the tube filled with air are 20 cm. apart. What is the wave-length and the frequency? The metal rod has a length of 60 cm. What is the wave-length in the metal produced by these vibrations? What is the velocity of sound in the metal?

10. A flute player warms his instrument up from 12° C. to 27° C. by playing on it. Assuming the air in the tube still to be as dry as usual, in what proportion have the frequencies of the notes given by the instrument increased? If the air contains more water vapor than before, does this increase or diminish the change?

11. An organ pipe produces 4 beats per second when sounded with a tuning fork of frequency 500. What two values might the frequency of the pipe have? If one could momentarily warm the pipe up a little, would this enable one to tell which of the two was correct? Explain.

12. A lighthouse sends simultaneous signals to a ship by sound waves in air and waves under (salt) water, at a temperature of 0° C. They are each heard on the ship, and arrive there 4 sec. apart. How far is the ship from the lighthouse?

CHAPTER 18

OTHER TOPICS IN SOUND

Pitch or frequency. Upper and lower limits. The term "pitch" as used by physicists is equivalent to frequency. Musical notes have a range in pitch from about 30 to 5000 per second. Below 25 or 30 the average ear hears separate puffs in the air, instead of a continuous sensation. Organ builders supply pipes in the largest organs whose fundamental tones go down to about 16 vibrations a second; it is now believed that we hear only the overtones from such pipes, and not the fundamental tone. Recent developments in phonographs have so greatly improved this method of sound reproduction that it is now possible to obtain records of ***pure*** tones at these low frequencies. With these tests can be made of the lower limit of hearing which are free from the confusion due to overtones, and lead to the result stated above.

Above about 5000 per second the sound becomes shrill and loses its musical character, though such high frequencies occur in music as upper overtones in certain instruments, such as the organ, the violin, and the piccolo. The average ear can hear to an upper limit somewhere between 18,000 and 22,000 per second, younger ears usually going higher than old ones. This higher range of frequencies (from 5000 up) is useful in giving character to noises and to musical tones, and especially in helping to form consonants in speech, but it is otherwise rather shrill and objectionable. Squeaks in machinery and hissing sounds come under this category. These high frequency vibrations may easily be produced

for the purpose of testing the upper limit of hearing by making a short steel rod vibrate either transversely or longitudinally. The transverse vibrations (p. 237) may be the louder ordinarily, but the longitudinal are better for this purpose. They are likely to last longer, and their frequencies being inversely proportional to the length of the rod, can readily be calculated from the known values for longer rods. If a long thick steel rod is held in the middle and tapped longitudinally, it will give out a high note which is persistent and the more easily heard the thicker the rod. A rod half as long will give a similar note of double the frequency, and so on. Thus a series of rods may be prepared, so as to carry the frequency through the upper limit of hearing into the region of "supersonic" vibrations. The same sort of test of the upper limit may be made, though less satisfactorily, with a "Galton pipe," an adjustable high-pitched whistle.

Supersonic vibrations. These are the inaudible vibrations above the upper limit of hearing. They have been produced and their frequencies have been measured up to and beyond five million a second, that is, into the region known as radio frequencies. They seem to be carried through the air at about the usual speed for sound waves, though carbon dioxide absorbs them, and if present in considerable amount in the air will affect their transmission over long distances. They can be concentrated by a curved mirror, such as is used in an automobile headlight, and thus directed toward a definite spot. They may almost equally well be produced under water, as submarine waves, and used as a means of communication between ships which could not be detected by anyone who is in a different direction, or who lacks the proper instruments.

The chief source of supersonic vibrations is a quartz plate which is made to oscillate in a direction perpendicular to its surface by means of the piezo-electric effect (p. 319). The mechanism for receiving these vibrations involves the same device.

Measurement of pitch. The frequency of a vibrating body may be measured with accuracy by several methods. Tuning forks can be obtained with known frequency and these are much used as standards of pitch. Other frequencies may be determined by comparison with them. If the pitch of a tuning fork itself is to be measured, it should be mounted horizontally with a light bristle attached to it. A smoked glass plate may then be drawn under it, so as to get a wavy line as a trace. Beside this must be run a time record, usually made electrically from a clock; so that by comparison of the two traces the number of

vibrations per second given by the fork may be counted directly. Figure 18–1 shows a record in which the time marks were put on by a swinging pendulum, whose period was separately determined.

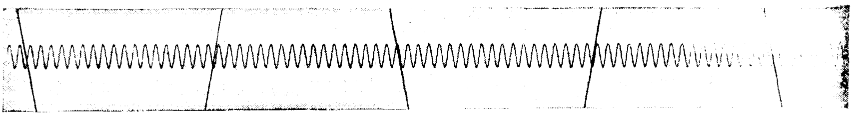


FIG. 18–1
Timing a tuning fork

A better way, called a *stroboscopic method,* is to view the fork, preferably enlarged by a lens, through a perforated disc which is rotated by a motor (Fig. 18–2). If the motor speed can be altered at will, and held steadily at any desired value, a speed can be found at which each brief glimpse of the fork obtained through one of the holes in the disc catches it at the same part of its oscillation; so that it appears to be at rest. If this condition is not quite secured, the glimpses will show the fork at progressively different points, so that it will appear to vibrate, but very slowly. At the correct speed, the revolutions of the motor may be recorded by a speed counter, and the number of holes passing by the eye per second found very precisely, and this is, of course, also the number of vibrations of the fork. This method may be applied to any form of vibrating body, or to revolving wheels with spokes, and the like.

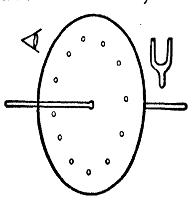


FIG. 18–2
A stroboscopic experiment

A stroboscopic effect may sometimes be seen in moving pictures which show the spokes of automobile wheels apparently at rest, or even rotating backward, while the car moves forward.

Comparison of pitches. Pitches may be accurately compared by several methods. If a phonodeik is available, the frequencies of any two tones may be compared when they are in any simple ratio to each other. If the two notes are sounded together, the wave pattern shown by the phonodeik will remain unchanged so long as the frequencies are exactly in a simple ratio. But if they differ by ever so little from that ratio this figure will change with time. Figure 18–3 shows records of the curves obtained in this way for two sounds in

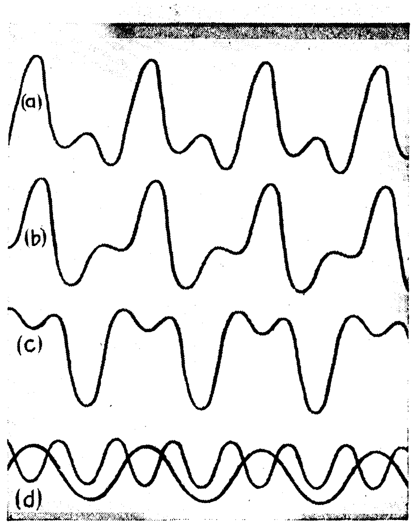


FIG. 18–3
Various forms of the complex vibration formed by two simple ones with frequencies in the ratio 1 : 2 and different phases

the frequency ratio 1 : 2. The curves (*a*), (*b*), and (*c*) are formed by two vibrations each of which keeps its amplitude unchanged throughout, the differences in the curves being due to changes in phase only. At the bottom (*d*) we see the two vibrations drawn out separately which when combined produce the curve (*c*). By watching these changes in wave form two forks, or pipes, may be tuned with almost unlimited precision in any simple ratio such as 1 : 2, 2 : 3, 3 : 4, 2 : 5, etc. Comparisons of pitch of this sort are far easier to carry out and more accurate than direct determinations.

Fork overtones. A tuning fork is supposed to give a single pure tone, but it does so only when struck gently with a fairly soft hammer at the proper place, the location of which can be found by trial. It has overtones which are much higher than the main note of the fork; they are not harmonic, and the first overtone usually has a frequency about six times that of the main tone. A light touch with a hard object will elicit them, different ones appearing for different points of impact. (N.B. A strong blow with a hard object will injure a good fork.) The nodes of each vibration may be found by pressing the fork with the tip of a lead pencil in different places, while tapping it, until a point is found where the presence of the pencil does not hinder the vibration.

The Doppler effect. If a source of sound is in motion toward an observer, or the observer in motion toward it, more waves are received per second by the observer than are sent out by the source. If the wave from the source travels a distance in one second equal to its velocity, V, and the *observer moves toward it* in one second a distance equal to his velocity u, the observer will in one second receive the number of waves f which he would have received at rest, plus the number he has overtaken by his speed, which will be the number of times the wave-length goes into his velocity. Hence the frequency f' received by the observer is $f + u/\lambda$. But $1/\lambda = f/V$. Hence $f' = f + fu/V$. The change in apparent frequency due to the observer's motion is fu/V. An equal change in the opposite direction occurs if the observer recedes from the source. When the *source is in motion* and the observer at rest, changes in frequency occur which are approximately the same, if the velocity is small compared with the velocity of sound. More accurately, the frequency change in this case is not exactly the same as before, and is not quite the same when the source is receding as when it is approaching. It can readily be calculated by examining the changes in wave-length produced by the motion of the source.

The law of change of frequency due to motion is known as Doppler's principle. An everyday illustration is the change in pitch which is heard as a locomotive rushes by with its bell ringing. A similar effect occurs with light waves.

What the ear can do. Our mechanism of hearing, limited as it is in range, is still remarkably successful. Sounds of moderately high frequency are especially keenly heard. It has been shown that a sound of frequency about 2000 vibrations per second can be heard if the particles of the air in the region of the ear are moving through a distance no greater than their own diameter, which we have seen (p. 146) to be about 10^{-8} cm. At the same time the ear can

discriminate between tones or musical notes whose frequency differs by less than one per cent; it can detect as separate two sharp sounds that occur less than a twentieth of a second apart, and it can analyze a complicated sound into its components by concentrating on one frequency at a time, or sort out a single sound from a confusing mixture. One might, for instance, be hearing at the same time the rustle of wind in the trees, the splash of waves against the shore of a lake, the songs of birds and the voice of a friend. The resulting mixture of tones and noises produces a complication in the form of the sound wave entering the ear which is

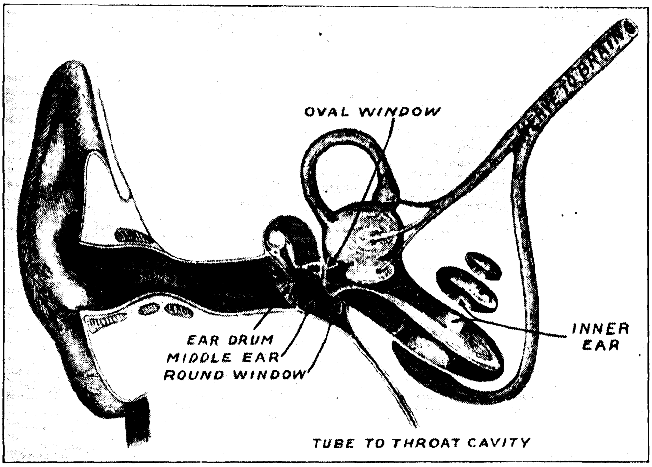


Fig. 18–4
The ear (after Czermak)

staggering to contemplate; yet, each of these sounds can with close attention be heard separately, very little disturbed by the presence of the others. Part of this ability comes from our having two ears, which enable us to determine the direction from which any sound comes, and to isolate it from others in this way. Part of the action must occur in the brain, rather than in the ear. Such difficult matters we gladly leave to the physiologist and the psychologist for solution. As the nerve impulses have electrical features, the process of hearing is probably, in part at least, an electrical one, and all the sciences will be needed to find its complete explanation.

Our hearing mechanism. Our hearing apparatus is divided into three chambers shown in Fig. 18–4. The first, the ***outer ear,*** consists of a short, narrow tube closed at its inner end by an elas-

tic diaphragm, the ***ear drum.*** The ***middle ear*** contains a curious arrangement of three little bones, linked together. One end of this system is attached to the ear drum, the other to a diaphragm, the ***oval window,*** which serves as one of two which close in the ***inner ear;*** the ***round window*** is the other. The purpose of the middle-ear mechanism seems to be to act as a lever, and give a larger force, with diminished amplitude of motion, in order to produce the vibrations in the inner ear. This is necessary since the inner ear is filled with a liquid, whose inertia makes it hard to move. The inner ear contains a coiled spiral enclosure with bony walls, shaped like a snail shell, (seen in section in Fig. 18–4), which is divided all along its twisted length, except at the tip, by the ***basilar membrane;*** along this membrane lie the auditory nerve endings. If the spiral were straightened out it would resemble

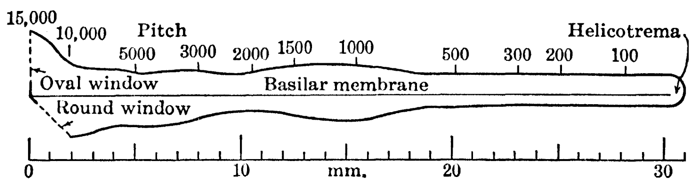


Fig. 18–5

The inner ear straightened out, and much simplified. The scale of mm. shows actual dimensions. The frequencies (above) are placed opposite to those parts of the basilar membrane which are most agitated by sounds of those pitches (after Fletcher).

Fig. 18–5. The vibrations are delivered by the middle-ear mechanism to the oval window, O. As the walls of the inner ear are rigid, and the liquid practically incompressible, something must give way when O moves, and the round window, being flexible, appears to be there for that purpose. The vibrations of the fluid then consist of back and forth surgings (alternating currents) from O by way of the opening at the upper end (the helicotrema) to R and back; except that the basilar membrane may yield to a certain extent, and thus transmit the vibrations from O to R across itself. The fastest oscillations appear to be transmitted almost entirely by the yielding of the lower part of the basilar membrane, the inertia of the liquid in the whole tube being too great to allow it to participate in so rapid a motion. For slow motions the alternating currents go by way of the open end, the stiffness of the membrane B being sufficient to prevent motion across it. For intermediate pitches both actions occur, but the yielding of the membrane is

the more important. Our ability to distinguish different pitches arises from the fact that at a particular pitch the energy of the vibrations passes more through one part of the basilar membrane than it does through any other. The nerve endings that lie in that part convey to the brain the impression of this pitch. Those that lie nearer the upper end are most agitated when the current goes that way (low pitches); those near O respond most to high pitches. Figure 18–5 shows by the scale of frequencies marked along the membrane where the response is greatest for each pitch.

This theory is not the only one that seeks to explain the action of the ear; but it is one to which few objections can be raised from a physical standpoint. It is based on an earlier theory of Helmholtz [1] and was developed by Fletcher and Wegel.[2] A full account of it is given in Dr. Fletcher's book already referred to.

Intensity and loudness. The intensity of a sound at a point may be defined as the amount of energy flowing per second through a square centimeter of area about this point perpendicular to the direction of the sound. This will depend on the amount of energy being sent out by the source itself, and on the distance from that source, as well as on the absorbing power of the surrounding objects. In free space the intensity will vary inversely as the square of the distance from the source (if the source is small), but it should be remarked that in practice "free space" is almost never obtained, so that this law of variation is not followed. In a partly enclosed space it usually happens that the variation with distance is less rapid, and rather irregular. The ***loudness*** of a sound is partly a psychological matter. It is difficult to measure, and until recently has been supposed to vary as the logarithm of the intensity; this relation is not exact. It also varies greatly with the pitch. It takes on the average some thousands of times more energy to produce an audible sound at a pitch of 32 per second than at 2000. The individual ear, moreover, is not very uniform in its sensitiveness. One's right ear is not usually exactly like the left, and two sounds of different pitch but equal intensity

[1] H. L. F. von Helmholtz (1821–1894), German surgeon, physiologist, physicist, mathematician and philosopher, a combination difficult for any but a genius, even in those days. He first clearly formulated the principle of the conservation of energy. His chief books are two great works on physiological optics and acoustics containing many original theories and inventions.

[2] H. Fletcher and R. L. Wegel of the Bell Telephone Laboratories, New York.

may not sound equally loud in each ear. Occasionally an ear is found which has a marked lack of sensitiveness for certain narrow ranges of frequency, perhaps quite unknown to its possessor. An electrical apparatus (the "audiometer") has recently been developed in the research laboratories of the Bell Telephone system by means of which a curve showing the sensitiveness of any individual ear for different pitches can be obtained, as in Fig. 18-6. This apparatus can produce any one of several pitches in a telephone, and by turning a dial one can make the note fade to inaudibility.

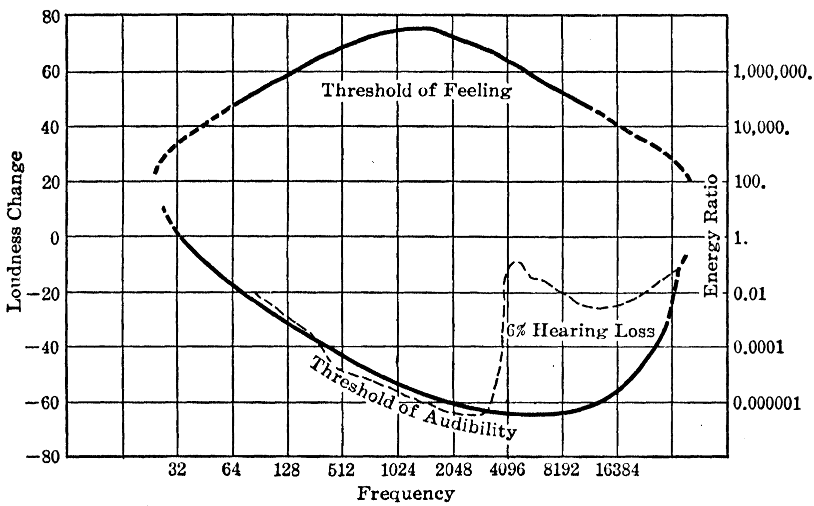


FIG. 18-6

Curve showing the energy needed at each frequency to produce the faintest audible sound; also (dotted) the curve for an individual ear which is slightly deaf for high pitches.

When this limit is just reached, the reading of the dial then indicates the sensitiveness of the ear at that pitch.

Resonance. Sympathetic vibration. There is an important mechanical principle which we have not yet considered. It has applications in the study of vibrations in all branches of physics. This is ***resonance,*** or, as it is sometimes called, sympathetic vibration. If one is agitating any body which has natural rates of vibration, for instance, the long rope whose harmonic vibrations were studied earlier (p. 236) and if ***the period of the agitation corresponds to one of the natural periods of the body,*** then the body responds vigorously, and this action is called resonance. A bridge is said to have been destroyed in England by soldiers

marching over it and stepping at a rate which coincided with the fundamental frequency of vibration of the bridge. A child in a swing can be made to move through a considerable distance by giving it gentle impulses at appropriate times. The impulses must have a definite period, and this must be the same as that of the swing; if it is, each push does work on the vibrating body, and stores more energy in it, until at last the friction involved in the motion becomes so great as to prevent any further growth of kinetic energy. Two tuning forks exactly alike, each mounted on a hollow box, may be placed near each other; then, if one fork is struck, the sound waves begin to agitate the other. This action is very feeble, it is true, but each impulse coming through the air is properly timed for resonance to occur. After a second or two one may stop the first fork, and the vibrations of the other will then be audible. The experiment works most successfully if the open ends of the two boxes face each other, as the vibrations are really transmitted not from fork to fork directly, but from one air chamber to the other.

If one holds down the loud pedal of a piano, thus freeing all the strings so that they can vibrate, a note sung briefly in the vicinity of the instrument excites those strings whose natural frequencies correspond, and the voice is returned as a faint, prolonged echo. Resonance in such a case occurs not only with the fundamental tone (of the voice, or of the string) but for each of the strong overtones; so that the quality of the voice will be somewhat reproduced, as well as its pitch.

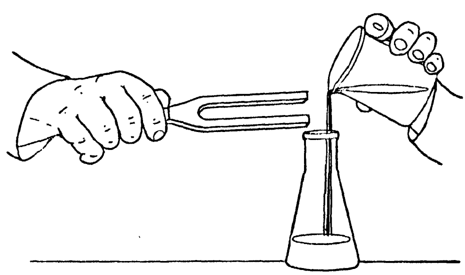


FIG. 18–7
A resonance experiment

A pretty example of resonance may be produced by holding a vibrating, unmounted tuning fork over a bottle (Fig. 18–7) into which water is being poured. When the water surface reaches a certain height, the response of the air column becomes very loud; taking away or adding a small amount of water destroys the agreement of frequency, and the response ceases. Other examples of resonance are seen in the undue response of the phonodeik (p. 262) or of a loud speaker to frequencies which are equal to those of the

natural vibrations of its diaphragm or horn. Important examples of resonance are found also in electric oscillations (pp. 417, 433), and even in connection with light waves.

Forced vibrations. A tuning fork is usually mounted on a hollow box whose air column is in resonance with the fork. What one hears from such an instrument is the air column, as can be verified by covering the open end of the box or by dismounting the fork and noting how little sound it makes by itself. If this fork is sounded in front of the air column of its own box, a loud sound is generated by resonance; but this does not occur if the box is one belonging to another fork, as the periods are no longer equal. If, on the other hand, the foot of any vibrating fork is touched to a table, the sound is again strongly heard, and this experiment succeeds equally well whatever the size or shape of the table. Any large light surface will do which can be shaken by the fork, and which can, in turn, shake a large body of air. The fork cannot do so alone, as it has too small a surface. Such a method of enlargement of the sound is quite different from resonance, and should not be confused with it. It is called ***forced vibration.*** It is by means of this action that a piano or a violin is able to make a loud sound at any pitch.

Musical instruments. Musical instruments have in general two parts, which we may call a *generator* of vibrations and an *amplifier*. The generator supplies the energy, and usually fixes the pitch, though the rest of the instrument may serve to keep it constant. The amplifier enlarges the sound by one of the two ways given above, resonance or forced vibration.

Stringed instruments. A stringed instrument has strings (or wires) which when plucked, bowed, or struck by a hammer acquire energy enough to shake the body of the instrument (as in violins, etc.), or a thin sheet of wood specially provided for this purpose, called the sounding board (as in the piano). What we hear comes from the large surface and not from the string direct. A "practice violin" is an instrument sometimes humanely given to beginners, which is barely audible; it consists of strings and a frame only, lacking the usual light body with its large surface. Resonance in a violin or piano would be very objectionable if it occurred within the range of the instrument's own frequencies; for, when the notes were sounded at which it occurred, the instrument would emit a volume of tone many times what it gave for other notes. Violinists sometimes meet with this phenomenon in poor instruments, when it is known as the "wolf note," presumably because it howls. Good instruments have their main resonant periods (if they have any) below the range of the instrument, and are equally responsive all over the scale of notes they are intended to produce. The amplification occurs by forced vibration only.

Different pitches are produced in stringed instruments by the use of wires of different length, thickness, and tension. The heaviest "strings" are wrapped with wire to make them heavy yet flexible.

As the standing waves in a wire are produced by ordinary waves traveling back and forth in it, one would expect the frequency to depend on the velocity of such waves. In the case of a string under tension the elasticity of the string is equal to this tension. Hence the velocity formula $V = \sqrt{e/d}$, (p. 243) becomes $V = \sqrt{T/m}$, where T is the tension in dynes and m is the mass in grams per centimeter. In the case of the simplest vibration (the fundamental of the string) the length of the string (L) is the distance from one node to the next, which is half a wave-length; or $\lambda = 2L$. Hence

since $V = f\lambda$ we have $$f = \frac{V}{\lambda} = \frac{\sqrt{\frac{T}{m}}}{2L}.$$

This formula shows how the frequency varies with the length, mass and tension of the string. It follows from this, for instance, that the tension must be made four times as great in order to raise the pitch one octave.

Wind instruments. Instruments which operate by means of air columns are of two main sorts, those in which nothing vibrates but air, and those which have a mechanical vibrator, consisting of a reed, or the player's lips. In these instruments the generator is either a thin stream of air or a reed, which supplies energy but usually has no very definite period of its own. The air column responds by resonance, emitting either its own fundamental tone, or one of its harmonics, or perhaps a group of several of these vibrations. A large body of air inside the tube is thus set into motion, communicating directly with the outer air, and sending out a good volume of tone. The inertia and the elasticity of the air column combine to make it vibrate at a constant rate, and thus, by reacting back on the generator, to keep the latter from varying in pitch. Different notes are obtained by selecting different harmonics of the air column or by altering its length. The latter change is accomplished by opening side holes, which virtually cut off the instrument there (as in flutes, clarinets, etc.), or by inserting additional lengths of tubing, either by valves (cornet, French horn, etc.) or by sliding tubes (slide trombone).

The human voice. This familiar instrument is one of the reed type, similar to one class just mentioned. The sound is generated by the vocal cords, whose frequency can be varied by changes in their tension, length and thickness. The starting and stopping of the vibration is governed by the flow of air which is under our control as we open and close the lips, and alter the size of the mouth and throat cavities. The tone quality is markedly affected by the shape of the latter cavities. Figure 18–8 shows the form they take for two familiar sounds.

Speech has been imitated by an instrument consisting of a bellows blowing air through a reed, which then passes out into the

open air through a wide flexible tube. If the tube is pinched at the opening (as in lip action), or varied in width at a point midway (as by the tongue), many simple syllables can be intelligibly uttered.

Consonant intervals. The ear, even when not especially "musical," recognizes a marked similarity between any tone and one whose frequency is one-half, or twice as great as its own. Thus musicians call such notes by the same letter or name (e.g., *C* or *do*). Any two notes whose frequencies are in the ratio 1 : 2 are said to be an octave apart. When sounded together they make a "consonant" interval, that is, an agreeable combination. Why

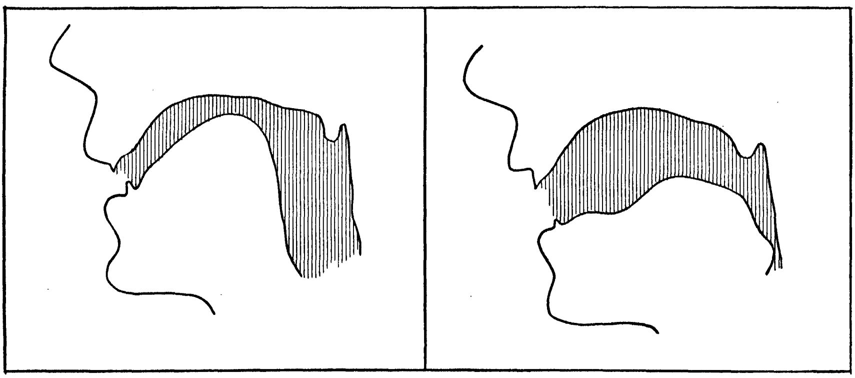


Fig. 18–8
X-ray sections of the mouth and throat cavity in uttering the sounds (left) *e*, as in "feet," and (right) *a*, as in "part."

they sound so much alike separately, and go so well together, are two effects which are probably due to the same cause; they occur together in the natural series of harmonics, and they make a combined motion in the air, and in the ear, which is ***simple.*** The form of the curve, as calculated, or as revealed by the phonodeik, is shown in Fig. 18–3 (p. 267) in several of its possible aspects. In Fig. 18–9 is shown the curve for two notes in the ratio 1 : 3, the relative strength of the upper note being progressively increased. The ear seems to enjoy simple combinations of periodic motions more than very complex ones, and ***simple curves arise from those combinations in which the ratio of the frequencies is a simple one.*** Musicians now recognize intervals as more or less consonant in the order of the simplicity of their ratios. The frequencies in the harmonic series are in proportion to the numbers 1, 2, 3, 4, 5, 6, etc., and any two or more of the lower members of this series are agreeable when sounded together. The interval 5 : 6

was not considered so in the early history of music, and 6 : 7, 7 : 8, etc. are not pleasant intervals to our ears now, as their motions are too complex. As we go farther up, such intervals become more ***dissonant,*** the maximum dissonance (in the middle of the scale) occurring when the curve form repeats itself in a cycle some 20 to 40 times a second. Such a curve, as in Fig. 17–14, (p. 259) represents a very dissonant effect, a musical half-tone, ratio 15 : 16.

The chief consonant intervals are as follows: octave, 1 : 2; fifth, 2 : 3; fourth, 3 : 4; major third, 4 : 5; minor third, 5 : 6; major sixth, 3 : 5; minor sixth, 5 : 8.

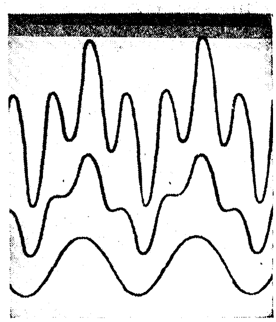


FIG. 18–9

Various forms of the complex vibration formed by two simple ones with frequencies in the ratio 1 : 3, the relative phases being kept constant, but the amplitude of the upper note being increased.

Musical scales. In early times a person who desired to express himself in a burst of song probably came to prefer in the long run to have at his disposal some sort of ***musical scale,*** after the fashion of a ladder; and after long use, one likes best those ladders whose steps are not very unequal. In forming a musical scale the consonant intervals cannot be entirely ignored, since in any but the most primitive music several tones will be sounded simultaneously. The attempt to make a scale having equal steps and including ***nothing*** but consonant intervals has met with only partial success. Table XXII shows part of such a scale, which we shall call the ***consonant diatonic scale.*** Most of the notes of this scale form exact and simple ratios with one another. Thus, $C' : F = 2 : \frac{4}{3} = 3 : 2$, a perfect fifth; and $G : D = \frac{3}{2} : \frac{9}{8} = 4 : 3$, a perfect fourth. But, if we try $A : D$, which should be a perfect fifth, we find it slightly out, and $F : D$ fails also. Thus not all the intervals among the notes are correct, and to cure this defect would require

TABLE XXII

Musical Note	C	D	E	F	G	A	B	C'
Frequency on the consonant diatonic scale	256	288	320	$341\frac{1}{3}$	384	$426\frac{2}{3}$	480	512
Ratio to C	1	$\frac{9}{8}$	$\frac{5}{4}$	$\frac{4}{3}$	$\frac{3}{2}$	$\frac{5}{3}$	$\frac{15}{8}$	2
Frequency on the scale of equal temperament	256	287.4	322.6	341.8	383.7	430.5	483.2	512

additional notes. An extra D, for instance, whose ratio with C was 10 : 9 would make the interval $A : D$ right; but it would be very confusing to have duplicates of each note to be used at times only. In fact, it can be shown that we should need at least four notes in place of each one now used in order to be able to play with this scale in every musical key, and have every interval exact. This would be quite unmanageable for most people, though it has been done.

The consonant diatonic scale contains unequal steps. Some of these, e.g., $E : D = 10 : 9$ and $D : C = 9 : 8$ are nearly the same and are called whole tones, while $F : E = 16 : 15$ is a half-tone, though it is not exactly half of either of the whole tones. By filling in the whole-tone intervals with additional notes (C sharp, etc.; not given in the Table) we get in all 12 half-tone intervals to the octave, which is the usual number; but on this scale these intervals cannot be exactly equal.

Scale of equal temperament. In practice, a compromise scale is actually used in all modern music, called the scale of equal temperament, with 12 intervals in an octave, all exactly equal, and each with a ratio equal to $\sqrt[12]{2}$, or 1.05946; that is, each note of the scale has a frequency 1.05946 times the next lower one. This scale gives no really perfect consonances. A piano-tuner must tune his instrument so as to avoid perfect intervals, and avoid them by a definite amount. The frequencies are shown in Table XXII. The fourths and fifths are fairly good; thirds and sixths are poor. The scale is the one in universal use because it is alike in all musical keys, and hence allows us to play modern music, in which frequent changes in key occur; and its defects are small enough so that none but the most sensitive are aware of them.

ARCHITECTURAL ACOUSTICS

The acoustics of buildings passed from what might be called the era of superstition, the era in which defective halls used to be strung with wires, to an exact science about the year 1900, and this advance was due to the labors of W. C. Sabine.[1] A very brief account, only, can be given here of this fascinating subject.

[1] The late Professor W. C. Sabine of Harvard University. His papers were collected after his death in 1919, and published as "Collected Papers on Acoustics," 1922 (Harvard University Press).

Absorption of sound energy in a room. Sound energy is like every other sort, in that the principle of the conservation of energy applies to it. If sound is produced in a room, it can cease only because it is absorbed, or because it escapes through open doors or windows. Deflecting it, or "breaking it up" (a term sometimes heard) can do nothing toward diminishing the sound unless it introduces increased absorption by increasing the number of reflections at which absorption occurs. Absorption means turning the regular air motions of the sound waves into the random motions of heat. This process does not occur in the air itself under ordinary conditions; it is best accomplished by porous bodies. Hair felt, plates of compressed vegetable fibers, or rough plaster loosely laid, succeed best in this action. A compact audience, with its mass of small spaces surrounded by soft surfaces, is an almost perfect absorbent.

Effect of the room on loudness and quality. In a room with hard, non-porous walls and no soft furnishings, a continuous source of sound builds up a very considerable amount of sound energy, that coming directly from the source being added to that which left it some time before, but which has not yet died out because so little has been lost by reflection from the walls. The sound of a musical instrument is thus actually louder in a hard room than in a soft one, where the direct sound is the only one of much strength. Another important fact is that the absorption by soft materials is usually more pronounced for high frequencies than for low, with the result that an instrument (piano, for example), which is rich in high harmonics, may sound "brilliant" (as the musicians say) in a hard room, due to the loudness of upper harmonics, and on the contrary poor and dull in a very soft one.

Time of reverberation and its calculation. If the source of sound in a room is suddenly cut off, the sound does not instantly stop, but "reverberates" and may be audible for a time of a few seconds' duration. If the source produces a sound whose intensity is a million times as great as that of the least audible sound, the time during which this particular sound remains audible in a room after it is shut off is called the ***time of reverberation*** of the room, t, and this can be measured with sufficient accuracy by a practiced observer. The sound, in traveling about in the room at the rate of 1100 feet a second, must very frequently strike the walls. The

TABLE XXIII

Absorption Coefficients for Sound

Material	Coefficient	Material	Coefficient
Open window	1.00	Varnished wood	0.03
Hair felt 1″	0.58	Plaster on wood lath	0.034
Acousti-Celotex BB. $1\frac{1}{4}$″	0.70	Glass	0.027
" " C. $\frac{3}{8}$″	0.30	Brick	0.025
Compact audience	0.96	Marble	0.01
Akoustolith tile	0.36	Adult person	4.7 sq. ft.
" plaster	0.29	Wooden seat per seat	0.15 to 0.20 sq. ft.
Other acoustical plasters up to	0.21	Seat cushion per seat	1.0 to 2.0 sq. ft.
Carpets	0.15 to 0.20		

smaller the volume, V, of the room is, the more often this will occur. Each time the sound is reflected some of the energy is turned into heat by being trapped in the pores of the wall material. Thus the energy of the sound diminishes, and the time of reverberation is less, in proportion to two factors, the frequency of the "collisions" and the absorbing ability of the materials of the walls and other surfaces of the room. Experiment and theory both show that

$$t = \frac{kV}{A}$$

where A is the total absorption of the room and k is a constant, whose value is 0.164 if the measurements are in meters, and 0.05 if they are in feet. The absorption A is obtained from a number of ***absorption coefficients.*** With reference to a room, a sound is entirely lost if it goes out of an open window. Thus an open window is taken as a perfect absorbent, like an absolutely black body for heat radiation. The coefficient of absorption of a material, a rug, for instance, is the ratio of the absorption produced by it to the absorption of an equal area of open window. The rug is said to be equivalent to a certain number of square meters (or feet) of open window. The absorption coefficients of the materials usually entering into the construction or use of a room have been measured, such as the various common wall materials, glass windows, wooden paneling, furniture, rugs, curtains, and even an audience per individual, or per unit area covered by it. These coefficients are given in Table XXIII for a sound of frequency 512. By multiplying the coefficient of each material by

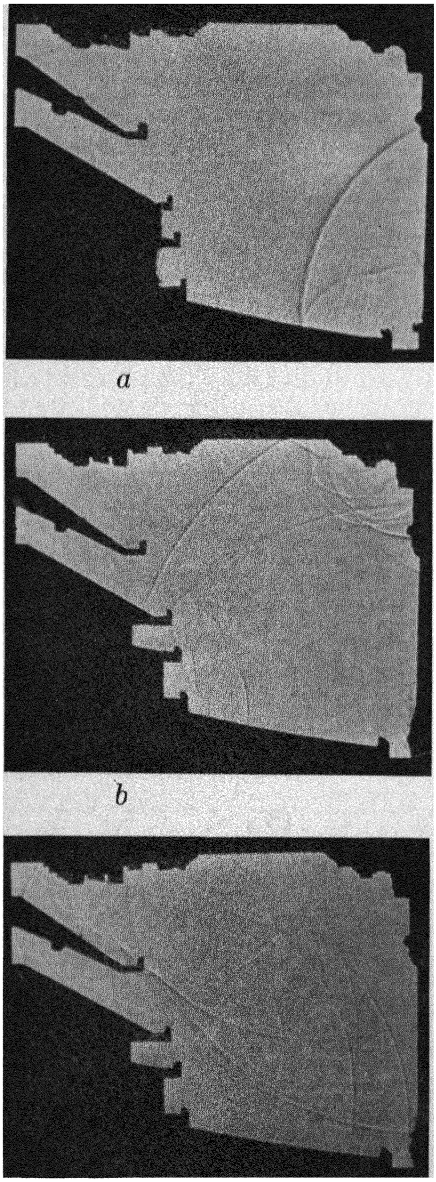
its area, one can, for instance, calculate the absorption of the whole audience, of the unoccupied chairs, of the furnishings, ceiling, walls, etc., and by adding these up, obtain the total absorption of the room, the quantity A in the formula. Thus the time of reverberation can be calculated, and this can be done as well before a building is constructed as afterwards. Sabine established the proper values for t under different conditions, which are from 2 or 2.5 seconds for a large hall used mainly for music, to 1 or 1.5 seconds for theaters of ordinary size, or for small music rooms, the best value varying in a linear manner with the cube root of the volume of the hall.[1] Hence the proper value for the total absorption of a hall can be calculated before it is built, and made correct by adjustments in the plans, usually in the choice of wall and ceiling coverings. Experience and theory both show that if the reverberation time is too long, confusion exists between sounds which should be separate, such as successive syllables in speech; if the time is too short, the room is too absorbent, and the volume of sound that can be created in it (as explained above) is disappointingly small, so that the room is said to be "dead"; musicians dislike to play in it, and a singer's voice fails to "fill" it.

There is now no more excuse for the construction of a hall with defective acoustics than for one with a leaky roof. Nevertheless, it does sometimes happen, and the usual defect is too long a reverberation time. Modern fireproof construction is often responsible for such cases, as the materials used are hard, and almost perfect sound reflectors. It is not difficult to calculate the amount of absorbent material that must be placed over the hard walls or ceiling of such a room to bring the time of reverberation down to a proper value. This material should not, as a rule, be placed near the source of sound, as it tends to deaden it unduly.

Study of acoustics by models. Sabine applied Foley's spark method (p. 244) to the study of the acoustics of buildings by making a plaster model of a hall, and using the "sound spark" as a source on the stage. The spreading wave then strikes all the objects in the hall in succession, and the course of the reflected waves can be followed. Figure 18–10 shows three stages in the growth of the wave in a model of a New York theater. In the

[1] The latter fact was established by F. R. Watson, whose book on the "Acoustics of Buildings," 1923 (Wiley & Sons) gives a good general account of the subject.

third stage the main wave has reached the farthest part of the hall, and the photograph shows an aggregation of reflected waves coming mainly from the ceiling, and causing an annoying echo to those seated in the front rows of the lower balconies. This concentration was later prevented by altering the form of the ceiling. The photographs give a vivid picture of the nature of reverberation.



a b c

Fig. 18–10

Sound deadening. The desire for quiet rooms is now frequently expressed, especially since psychologists have proved that more efficient work can be done under these conditions. Offices in which typewriter noises are prominent are now being treated; here the offending sounds are mainly of high frequency, and the materials used must be such as absorb these frequencies. Hospital rooms are sometimes made sound absorbent, so that street noises which enter the room will be deadened. Quiet restaurants near busy streets are now, if rare, at least a possibility. In such problems we are often especially concerned with the passage of sound through walls.

Sound transmission. Sound transmission through a wall is most often accomplished by the shaking of the wall as a whole by the sound waves. This passes the agitation through to the air

on the far side. This occurs particularly easily if the wall happens to have any resonant periods corresponding to the frequencies that are present. Sound also gets surprisingly well through cracks in plaster walls or around supposedly tight doors. Sound deadening materials for use in wall construction are usually efficient in proportion to their inertia; the more mass there is in a wall the less will it be moved by a sound wave, and the less energy will be passed on, or "get through." Sound often travels through and along the rigid framework of buildings in elastic waves which carry it surprisingly well; such sounds may arise from heavy moving objects (machinery in the basement, for instance) and can be reduced only by mounting the offending sources on some support more or less independent of the building structure. Sound proofing is still a difficult art, in which a wide knowledge of physics is of great advantage.

PROBLEMS

1. A disc fitted with 32 holes revolves at the rate of 720 revolutions per minute. Through it a vibrating tuning fork is seen ("stroboscopically") as though stationary. What is the frequency of this fork?

2. A moving picture taken at the rate of 16 photographs per second shows an automobile wheel (with 12 spokes) apparently not revolving though the car moves forward. Explain, and show at what speeds the wheel might really have been revolving.

3. Find the difference in pitch between an automobile horn vibrating at a frequency of 400 per second and its echo from a vertical cliff which the car is approaching at the rate of 30 ft./sec., the listener being in the car.

4. When an engine goes by, ringing a bell, we hear a change in pitch. Draw a curve illustrating (qualitatively at least) this change, plotting pitch against time.

5. Point out the fallacies involved in the idea that wires strung across the room will deaden sound in it. (Note that the absorption coefficient of metals is comparable with that of glass.) What will happen to the sound energy if the wires have resonant frequencies equal to those of the sound in the room? (What becomes of the energy of a vibrating wire?)

6. A hall has a volume of 100,000 ft.^{3} and a total absorption equivalent to 1200 ft.^{2} of open window. What is the effect on the reverberation time if an audience fills it, thereby adding to the absorption by 2800 ft.^{2}?

7. An experimental sound room has a volume of 1000 ft.^{3} and a reverberation time of 10 sec. If 40 ft.^{2} of an absorbent material are brought in and

reduce the time of reverberation to 2 sec., what is the coefficient of absorption of the material?

8. A hall has a volume of 180,000 ft.^{3} and an absorption due to the following items: an audience of 1000 persons; plaster, glass, and varnished wood surfaces in walls, ceiling, and floor 25,000 ft.^{2} (average coefficient 0.03); other items equivalent to 100 ft.^{2} of open window. Find the reverberation time and the amount of additional absorption that must be introduced to bring this time down to 1.5 seconds.

Additional Books on Sound:

E. H. Barton, "A Text-book on Sound," (Macmillan).
E. G. Richardson, "Sound," 1927 (Longmans, Green).
E. G. Richardson, "The Acoustics of Musical Instruments, 1929, (Edward Arnold and Co.).

CHAPTER 20

ELECTROSTATICS

Simple electric phenomena, 303; unit charge, law of force, 304; electric fields, 304; the electrostatic series, 305; conductors and insulators, the electroscope, 306; atomic structure, the electron, 307; the oil-drop experiment, 308; electrostatic induction, 310; charging by induction, 311; electrostatic screening, 311; ionization of the air, 311; ionization by collision, the electric spark, 312; ionization by flames, 313; effects due to moisture, 313; discharges from points, 313; lightning protection, 315; machines for generating electric charge, 315; potential difference, units, the volt, 316; condensers, 317; dielectric constant, Leyden jar, 318; the "breakdown" of a dielectric, electrostriction, piezoelectricity, 319; capacity, 320; displacement currents in condensers, 322; hydraulic analogue of condenser action, 322; energy of a charged condenser, 323; condensers in series and parallel, 323; atmospheric electricity, 324; electrometer, 325.

Simple electric phenomena. Back in the days of ancient Greece it was known that amber (which they called "electron") when rubbed would attract light objects such as thin shavings of wood. Later on, especially through Gilbert's experiments, it became evident that a large number of other substances showed similar effects. Contact or friction is necessary to develop them. The most convenient substances to use nowadays for such experiments are hard rubber (e.g., a fountain pen or a rod), glass, wool or fur, and silk. It is interesting to have also two small light balls (of pith, for instance) suspended by clean silk threads of a length of a foot or so; the balls should be coated with gold or aluminum foil, or slightly moistened with a salt solution such as calcium chloride, to make them electrical conductors.

If a suspended ball is touched by a hard-rubber rod which has previously been vigorously rubbed against a woolen coat sleeve, it will be repelled by the rod for a long time afterward. If another such ball is touched by a clean glass rod which has been rubbed with silk, it will be repelled by the glass rod, but attracted by the hard-rubber one; and the two balls if brought near together will attract each other. We say that the rods and balls are "electrically charged" or that a charge of electricity exists on them which is

static (i.e., at rest).[1] It is evident that there are two sorts of electrical charge, which we distinguish by calling them positive and negative, because they tend to annul each other if added together. Thus the two charged balls above will destroy each other's charges more or less completely if allowed to touch. The charge residing on the hard rubber when rubbed against wool is called negative, following a suggestion of Benjamin Franklin[2]; that on the glass, when rubbed with silk, is called positive. (If glass is rubbed with fur, a negative charge may be found on it.) Summarizing these observations, we have the rules that ***like charges repel each other*** and ***unlike charges attract.*** These effects remind one of similar ones in magnetism, and the similarity extends even further. There are, however, many differences, the most striking being the fact that charges can move readily, and independently of one another.

Unit charge. Law of force. An ***electrostatic unit of charge*** is defined, (like the unit magnetic pole), as one which repels a similar one at 1 cm. distance with a force of 1 dyne. The force between two small balls which are charged with e and e' units respectively is $F = ee'/d^2$ (Coulomb's law for electric charges), if d is the distance between the centers of the balls. The formula is strictly true if the experiment is done in a vacuum, though practically the same result is obtained in air. If it is tried in other media (oil, for instance), the force is diminished by a constant factor which is treated later (p. 318). Experiments in other media than air are difficult.

Thus we see that the law of force for electrical charges is similar to that for magnetic poles.

Electric fields. The region surrounding electrically charged bodies is called an electric field, and may be regarded as being

[1] We must be careful to recognize that we have learned nothing about these phenomena just by giving them names. Familiarity with electricity breeds a quite unjustified feeling of certainty that a charge is like a fluid which we can put on, or wipe off a body. This conception is useful as an analogy, but it may not be right.

[2] Benjamin Franklin, 1706–1790; American statesman and scientist. He first drew down electric sparks from the sky by means of a kite, discovered the nature of lightning, and devised a method of lightning protection. The stove still known by his name combines efficiency with the pleasures of an open fire. He suggested water-tight compartments for ships, made the first bifocal eyeglasses, and brought his many-sided genius to bear on a host of other practical problems.

filled with lines of electric force, endowed with tension and mutual repulsion, as Faraday supposed the magnetic lines to be. These lines pass from positive charges to negative ones, following the direction which would be taken by an isolated positive charge. We imagine a difference between the electric and magnetic lines inside bodies. Since electric charges reside on the surfaces of conducting bodies (p. 311), there are no charges inside them, and the lines must end on their surfaces; while in the case of magnetic lines, we imagine them to pass through the interior of magnets. In other words, the interior of a magnet is magnetized, but the interior of a charged conducting body is not charged.

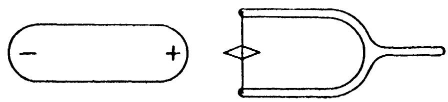


FIG. 20–1

Lines of electric force may be traced in the air near a highly charged body by means of an indicator in the form of a thin paper arrow supported on a stretched silk fiber; the arrow tends to point along the lines of electric force, and will do so if the thread is so thin that its own stiffness does not control the position of the arrow. Such a device is shown in Fig. 20–1, where the frame, which also serves as handle, is of glass. Another way of exhibiting electric fields is by the use of short bits of hair, such as can be obtained by trimming off a paint-brush, dusted on glass plates, just as iron filings are used for magnetic fields. Figure 20–2 was obtained in this way.

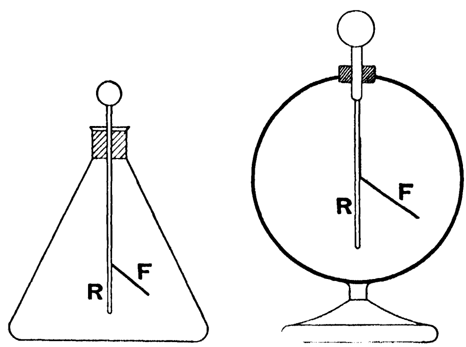
An electric field has an ***intensity*** at every point defined in a manner analogous to that of a magnetic field, as ***the force acting on a unit charge*** on a small body placed at the point in question. If E is the intensity of the electric field at a point, and e the quantity of charge placed there, the force acting on the charge is eE dynes.

The lines of electric force may be drawn in such numbers as to represent at every point the intensity of the field there, as was done with magnetic fields. Then, drawing a sphere around a unit point charge, with 1 cm. radius, we have a surface of 4π cm.^{2}, at all parts of which there is unit intensity. Hence it follows that ***there are 4π lines radiating from each unit of charge.***

The electrostatic series. By rubbing one substance against another and testing the charge on each, it is found that all substances can be arranged in a series such that any one becomes positively charged when rubbed against a substance below it in

largely on this factor. When electrically charged, the metal conducts this charge to all parts of itself, and the foil *F* and the part *R* of the rod are repelled by the like charges on them; but, of course, only the foil is light and flexible enough to move. It stands out at an angle which depends on the amount of charge, and serves as a convenient indicator of it. If *F* is viewed through a microscope furnished with a micrometer scale, this instrument may be used for measurement, rather than mere observation.

The relative ability of different substances to conduct electricity may be roughly tested by connecting a long thread or wire made of the substance to the metal knob on the top of the instrument, and bringing electrically charged bodies near the far end of this wire. For instance, a common cotton thread when wet enables one thus to affect the electroscope from a long distance, but when dry it acts as a nonconductor, or insulator, and does nothing. More exact methods of finding the conductivity of substances are given below (p. 342).



a FIG. 20-3 *b*
Two forms of electroscope

Atomic structure. The electron. Before proceeding further with the facts of electrostatics, it is convenient to consider the modern explanation of them. This is so simple that one who knows it may reason out in advance the result of many experiments which would otherwise seem obscure, and thus save much useless memorizing of isolated facts.

All matter is made up of ***atoms,*** usually in definitely constructed groups called ***molecules.*** It is now believed, on good experimental evidence, that each atom consists of two quite different constituents, a relatively small ***nucleus*** bearing a positive charge of a definite amount, different for each sort of atom, surrounded by negative electricity in equal amount; so that the whole atom, observed from any distance much greater than its own size, appears electrically neutral. The negative electricity occurs only in definite amounts, called ***electrons,*** which we might describe as the atoms of negative electricity in the original sense of this word,

meaning that they cannot be subdivided. The positive electricity in the nucleus must contain atoms of positive electricity, since the positive and the negative always balance, but there is an important difference in the way in which the positive and negative charges occur. The positive nucleus, though excessively minute, ***contains practically all the mass*** of the atom, while the electrons can be observed in some experiments described below (p. 441) as light electrical bodies with a mass which is measurable, but very small compared with that of the nucleus. The external part of the atom is entirely negative, and we shall see that all actions involving contact with other atoms (e.g., all chemical actions, or the emission of light by the atoms) concern the external structure only (i.e. the electrons). Thus, if two sorts of material are rubbed together, one may look at the affair crudely as a rubbing of the outer parts of atoms, which are made of electrons more or less tightly bound. If one material holds its electrons more loosely than the other, it may easily happen that some of them may be scraped off, and carried away by the material whose electrons are more tightly held. The result will be that the latter will show a negative charge, and the former an ***equal*** positive charge, due to the absence of its usual supply of electrons. In fact ***whenever any charge is produced, an equal and opposite one must be produced at the same time.*** That this really happens is readily shown by a simple experiment (p. 327, problem 18).

This theory also implies that whenever electricity flows through a solid conductor, it is the negative electricity that moves; the positive remains fixed with the atoms. In a liquid or a gas, where the atoms also may move, the mechanism may be different. We may also conclude that a non-conductor must be a body whose electrons are too tightly fastened to be free to move away from their natural positions.

The oil-drop experiment. The existence of the electron is so universally accepted and so fundamental that it will be worth while to consider one of the simplest experiments (though not the earliest) upon which this concept is based, which was carried out in 1909 by Millikan.[1] (See also p. 441.) Fine drops of liquid when

[1] R. A. Millikan, at that time Professor of Physics in the University of Chicago, now in the California Institute of Technology; one of America's foremost physicists; winner of the Nobel Prize in Physics in 1923. His book entitled "The Electron" gives an excellent account of this field of research.

shot out of the nozzle of a spraying device (or "atomizer") are likely to be electrically charged. If one of them is allowed to fall into a space where a vertical electric field exists, it may be possible to hold it from falling by the electric force which then acts on it. In such a case the force depends on the charge, e, and on the intensity of the electric field, E; it is equal to eE (p. 305). If the drop is just balanced, this force is equal and opposite to the weight, W, of the drop. The intensity of the field can easily be found (p. 321). The weight can be determined from a definite relation which exists between weight, diameter, and rate of fall in air, since the friction of the air holds back the lighter and smaller drops in a quite regular manner (see p. 48). Hence, if $eE = W$, and E and W are known, the quantity of charge e borne by the drop may be found. This was done for a large variety of drops bearing different charges, and ***the charge was found always to be a multiple of a certain small quantity of charge,*** whose value is 4.77×10^{-10} "E–S" (electrostatic) units. The arithmetical problem of finding this quantity is similar to the one of finding the weight of a single egg, given the weights of a large number of paper bags each containing a different and unknown number of eggs.

An important feature of the experiment was the possibility of increasing or decreasing the charge borne by the drop, while it was under observation. This is done by ionizing the air (p. 311) by means of X-rays (p. 582). If the drop is nearly balanced, rising or falling slowly, its rate of motion is suddenly changed at the instant when the drop catches another electron; and its new charge may again be measured. Thus Millikan was able to alter the charge on the drop at will, keeping the same drop under observation for many hours at a time. The results of this experiment showed that drops of liquid (oil, glycerine, mercury, etc.) whether they are electrical conductors or not, and whether they become charged by friction or otherwise, always carry, (or lack), an integral number of electrons, and change their charges by a small (whole) number of electrons at a time, usually only one. This far-reaching result indicates that the electron is the smallest quantity of electricity that can exist, and that it is found alike in all sorts of matter.

Figure 20–4 shows the chamber C into which the atomizer blew a cloud of drops, and the small hole H which was opened to allow a drop to fall into the observing chamber, where it could be

brightly illuminated from the side and viewed with a strong microscope; also the metal plates *P* between which the electric field could be created, by connecting them to the terminals of a high-voltage battery.

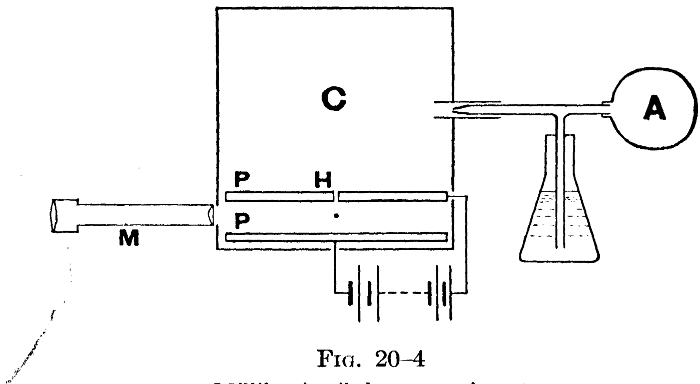


FIG. 20-4
Millikan's oil-drop experiment

Electrostatic induction. If a negatively charged hard-rubber rod is placed near a conducting body, the electrons in the conductor are repelled and are found in excess in a region as far away from the rod as possible, leaving a positive charge near the rod, as in Fig. 20-5*a*. If the conductor is an electroscope with the charged rod

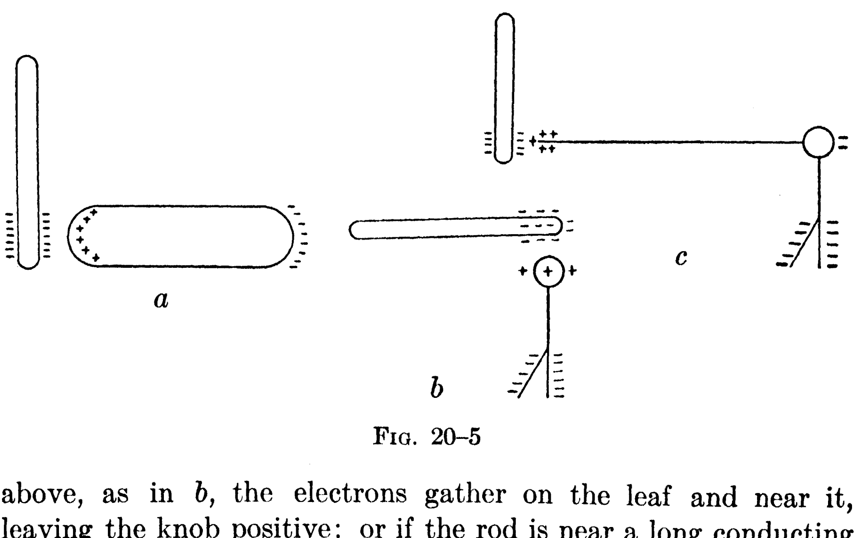


FIG. 20-5

above, as in *b*, the electrons gather on the leaf and near it, leaving the knob positive; or if the rod is near a long conducting wire connected to the electroscope, as in *c*, the whole electroscope will be negatively charged.

Charging by induction. If a negatively charged rod is brought near a pair of conductors, *A* and *B*, Fig. 20–6, each mounted on insulated supports, and placed so as to be in contact with each other, some electrons will be driven to the far side of *B* leaving *A* positively charged. If now *A* and *B* are separated a little and then the rod is removed, the extra electrons on *B* will be unable to return to *A* and both *A* and *B* will be (more or less) permanently charged; that is, they will keep their charges until they leak off into the air or along the not quite perfectly insulating supports.

In this process the charges on *A* and *B* have been created without any alteration of the original charge on the rod. This looks as though something had been obtained for nothing, in violation of some of nature's laws. It is not so, however; *A* and *B* attract each other when they are charged, and it takes some work to separate them. Thus energy is supplied from outside, which is stored as potential energy in the charges.

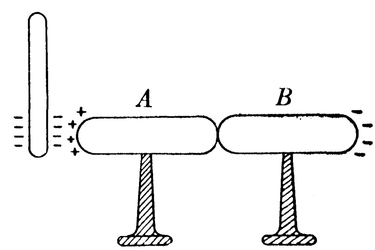


Fig. 20–6

Electrostatic screening. If it is desired to screen a body (an electroscope, perhaps) from outside effects, it may be surrounded by a shell of metal, or even of wire gauze. The charges induced on the screen will always be such as to annul any inside effects produced by the inducing charges outside. The outer surface of any conducting body in like manner screens the inside from electrostatic effects.

Ionization of the air. One cannot proceed far with experiments on electrostatics without observing the phenomenon of the electric spark in air. The crackling sound made by a comb in one's hair on a cold winter's day is an example, as are also the tiny sparks that are heard when one runs a charged hard-rubber rod over one's ear. The air is usually a non-conductor of electricity; its molecules are complete in themselves, and neutral, taking no observable part in electric actions so long as they remain so. But it is not difficult to create conditions in which disturbances occur of sufficient violence to break off an occasional electron from one of these molecules. When this happens, the electron usually attaches itself very soon to a neutral molecule, thus making it into a negatively charged particle which we call a "negative ion" (an "ion" is a

traveler); while the molecule it came from is a "positive ion." Ions are thus charged molecules (or sometimes atoms) carrying with them a quantity of negative charge (in the form of electrons) which does not balance the positive charges borne by their nuclei. They usually lack one electron and are then positive, or carry one more than usual which makes them negative; though often there may be an excess or deficiency of two or more electrons.

There are causes acting to produce a few ions in the air continually (Chap. 39). Even the most elaborately insulated electroscope will lose its charge gradually in this way; if it is negatively charged, for instance, it will attract positive ions, giving each of them the electrons they lack and sending them off as neutral molecules once more, but losing a little of its charge in the process, so that in time it loses all.

Ionization by collision. The electric spark. If an ion is placed in an electric field it experiences a considerable electric force which in so light a body soon creates a high velocity. If the field is high enough, this velocity may become so great that the ion can succeed in breaking off an electron from the next molecule with which it collides. If so, this creates two new ions, which rush off in opposite directions, urged by the electric field, and soon create four more, and these eight more, and so on. Remembering the activity and enormous numbers of the molecules in a gas, it is not hard to believe that a very large number of ions is thus almost immediately created, with the result that the air becomes a conductor and neutralizes the charges which were creating the electric field. It is an amazing fact that this whole process takes only a millionth of a second or so, and constitutes an ***electric spark.*** (Fig. 20–7). The energy of the electric field is dissipated as heat, sound, and light in the gas where the spark occurs.

Fig. 20–7
A four-inch spark in air

If we bring a hard-rubber rod, highly charged, nearer and nearer to a large conductor (one's finger, for instance), the electric field between the two increases as the charge approaches the opposite

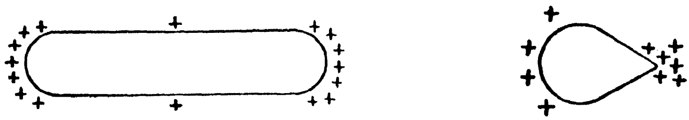
induced one which it creates in the conductor, and at a certain distance a spark occurs, which discharges that part of the rod, but not all of it. Since the rod is a very poor conductor (if it is clean), the excess electrons elsewhere on it cannot flow along the rod to the spark and so pass away. The ionized air in the region of the spark reaches over only a short distance, because the positive and negative ions in the air attract each other and in the course of a short time (of the order of one second) they have all recombined and the air has become normal and non-conducting once more.

Ionization by flames. The chemical activity in a flame produces very rapid motions of the gas particles taking part in the reactions due to the heat developed, so that a few of these are moving fast enough to produce ionization by collision. Ions are also produced by the chemical actions themselves. Thus it happens that flames are conductors of electricity, and since it takes a little time for all the ions to recombine, the warm vapor rising above a flame is also somewhat ionized and hence conducting. If one wants to discharge a charged rod, or a bottle which has been rubbed along a dry table, or any such object, one has only to put it in the stream of rising air a foot or more above a flame and the charge disappears as if by magic. This is useful in carrying out experiments on electrostatics, especially in heated rooms in winter, as almost everything becomes charged even by being moved, and unexpected and puzzling effects often occur.

Effects due to moisture. In moist air, an invisible film of water condenses on the surfaces of practically everything, and as this is never quite free from dissolved impurities which make it electrically conducting, no insulators act well under these conditions. Experiments on electrostatics are therefore best carried on in winter, when the humidity is low. Even then one's shoes, one's body, the furniture in the room, and the floor are all fairly good conductors, so that touching a charged body is equivalent to connecting it with the earth, which being so huge an object absorbs any charges we put into it as though it were infinite in capacity. The experimenter who from forgetfulness or otherwise happens to wear rubbers while he is working will find himself insulated from the earth (unless he is sitting down) and probably also rather highly charged. Occasionally a thick carpet is so dry that when a man walks over it he is likewise insulated from the earth, whereupon the friction of shoe leather against the carpet charges him highly, so that he can light a gas jet by the spark from his finger. But this experiment does not succeed in moist air.

Discharges from points. If a conducting body, some of whose surfaces are curved, as in Fig. 20–8*a*, is charged, most of the charge

is found on the ends, while very little resides on the straight surface. The shape shown in Fig. 20–8*b*, gives a marked concentration on the sharp end. This may be tested by touching a small coin mounted on an insulating handle to the spot under examination, taking off a sample, as it were, and transferring this charge to an electroscope, to see how large it is. If the end of (*a*) or (*b*) is fitted with a sharp projecting needle, the electric field at its tip becomes



a FIG. 20–8 *b*

so strong that ionization by collision occurs, and an electric discharge into the air carries off the charge on the body as fast as it can be put on. The ions repelled from the end drag the air with them and create an "electric wind" which is capable of blowing a candle flame or a rising column of smoke visibly out of shape. In the dark a faint violet glow (or "brush discharge") may be seen at the point (Fig. 20–9) emitted by the ionized nitrogen molecules in the air. This sort of thing is sometimes seen at the tips of flagpoles on high mountains, or the masts of ships, when it is known as "St. Elmo's fire." Similar brush discharges may occur along high-tension electrical transmission lines and are then called "corona" discharges (p. 426).

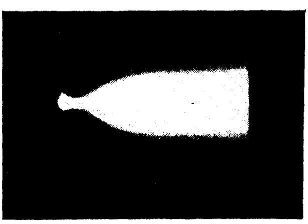


FIG. 20–9

A glow discharge between a point and a plane, photographed by its own light

A simple explanation of the concentration of the electric field on sharply curved surfaces can be given in terms of the lines of force, if they are imagined to be endowed with the two properties suggested by Faraday, namely a tension along their length and a mutual repulsion. On any conducting surface the lines of electric force leave the surface perpendicularly when the charge is at rest; for if it were not so, there would be a force component acting on the charge to which the inclined line is attached, which would pull the charge along with it. If the surface is plane, the lines of force must be few because they can exert their repulsion upon one another with maximum effect, since this is assumed to act per-

pendicularly to the lines. But on a curved surface (as in Fig. 20–10) near-by lines are inclined to one another, so that the repulsive force does not all act perpendicularly to both lines, and hence is not so effective. Thus more lines of force can end on the curved surface, and it follows that the concentration of charge there may rise to an extremely high value if the point is a very sharp one.

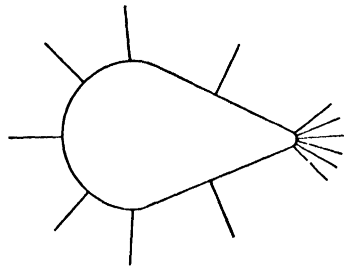


FIG. 20–10
Electric lines of force around a charged conducting body

Lightning protection. If a cloud is charged, say positively, it induces a negative charge on the nearer parts of the ground (Fig. 20–11) which may rise to a considerable value on a small elevation like a building. If the building is provided with metal rods well connected, electrically, to the ground, and these rods project into the air, ending in *very sharp points*, there will arise in times of danger a glow discharge at these points. This is a sign that ions are busily carrying the charge on the building off into the air, where it will be diffused and carried away by the wind, and thus reduce the chance of a lightning stroke on the building. Evidently, this protection will not be secured if the points are broken off, or blunted.

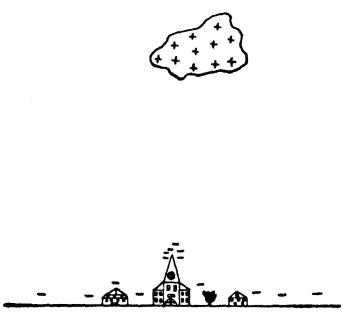


FIG. 20–11

This does not complete the study of lightning protection. When a lightning stroke occurs there are violent surges of electricity in all near-by conductors. These currents are more effectively provided for and a house is better protected if the rods form a sort of very open mesh or cage completely enclosing it. When this is done, and all vertical rods are well grounded, the protection is effective.

The large oil companies have tank "farms" where huge amounts of oil may be stored. Efficient lightning protection in such places is absolutely necessary. It is equally important for long-distance electric transmission lines.

Machines for generating electric charge. The process of charging by induction, as shown by the experiment with the rod and two insulated conductors (Fig. 20–6), serves as a basis for a machine called the ***electrophorus*** which is sometimes used for generating static charges. In this instrument a non-conducting base (of hard-rubber or wax) has a flat upper surface which can be charged negatively by rubbing it with fur. On it is placed a metal plate with an insulating handle. Even when the metal plate rests

on the base, it makes actual contact at only a few points, since neither surface is perfectly flat. As the base is a non-conductor, its charge cannot flow off through these points. The area of these points of contact is an insignificant part of the whole area, so that we may regard the two surfaces as actually standing a little apart, as shown in an exaggerated way in Fig. 20–12. The induced negative charge on the plate is repelled by the negative charge on the base, and escapes if the plate is touched with the finger. After that the plate may be removed, and the positive charge on it will be large enough so that a small spark may be drawn from it. The process may then be repeated indefinitely, and more and more positive charge obtained, without in the least diminishing the original charge on the base.

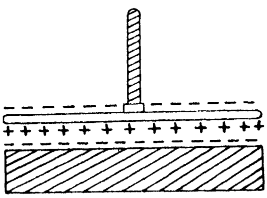


Fig. 20–12
The electrophorus

A machine for generating these charges would be mechanically better if it worked by means of rotation. There are several types of machine of this sort, of which the Wimshurst machine is one of the best. In these machines there are several conductors of tin foil mounted on a rotating disc made of an insulating material. Each conductor in the course of its motion first finds itself opposite to a charged body, while a "collector" draws off the repelled charge from its back surface. Then it moves on and passes another collector to which it gives its charge; after which it is ready to repeat the cycle. The collectors are metal combs with sharp points. The ionization produced at the points makes the air so good a conductor that it establishes effective connection between the tin foil and the collector. The collectors are each connected to a Leyden jar (p. 318) and to one knob of the machine. A large charge is built up in the jars, and is then capable of making a loud spark between the knobs.

Potential difference. Units. The volt. Water flows from one place to another if the first is higher than the second, and the rate of flow depends in general on the difference in level between the two places. If for some reason this difference in level could not be observed directly, it could be inferred from the work done, or the change of energy involved in the flow of the water.

In the flow of electricity there is something akin to the difference of level, which we call difference of potential ("P. D."). Since

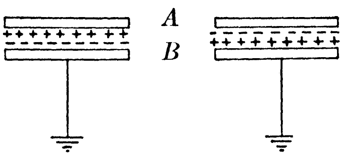
no electrical levels can be seen, we must infer them all from the energy changes involved. Two charged bodies have no difference of potential if there is no flow of electricity (current) when they are connected by a conducting wire. If there is a flow, the difference of potential is measured by the work done in transferring one unit of electricity through it. Thus ***the electrostatic*** ("E–S") ***unit of potential difference between two points is such that one erg of work is required to transfer one E–S unit of electricity from one to the other.*** This way of determining potential difference is useful in electrostatic problems but seems somewhat awkward. It is not easy to measure out one E–S unit of electricity, nor to find the work involved in transferring it from place to place. We shall see later that in practice we have easier ways of measuring potential difference and that the unit we all use is the ***volt.*** Approximately 300 volts make one E–S unit. This may be taken as one definition of the volt, though a more convenient one is given later, (p. 353).

For purposes of measurement the earth is taken as having zero potential, and all differences are considered with respect to it.[1] Those bodies from which positive electricity would flow (if it could) to the earth are said to have a positive potential; actually, of course, positive electricity cannot flow in a solid, but in this case electrons flowing from the earth into positive bodies are equivalent to a positive flow into the earth. Though the earth is taken to be at zero potential, there is good reason to believe that it is charged, and this charge may undergo fluctuations. As the earth is so huge that nothing we can do can appreciably affect its charge, we cannot find any steadier or more convenient standard of potential. We are reminded of the use of mean sea level as a standard of height, which in view of probable elevation or subsidence of the coast line with time is not ideal as a standard, but is the best we can get.

Condensers. If two flat metal plates are brought near together, one of them, B, connected to the earth, the other, A, charged, a state of affairs like that of Figs. 20–13a or b is reached. In a, while the positively charged plate A was being brought up to B, electrons flowed up from the earth, which can always be regarded as an infinite reservoir of them, to the surface of B, so as to get as close as possible to the attracting positive charge. In case b, electrons

[1] For calculations it is usual to consider the potential at an infinite distance from any charged bodies as zero.

were driven out of the upper surface of B to the earth. In each case the upper surface of A has very little charge, especially when A and B are very close together. The charges then become "condensed" on the adjacent surfaces and the pair of plates is called a ***condenser.*** A large charge can be placed in a condenser for the reason that the repulsion between one part of the concentrated charge on a plate and the rest of it is partially overbalanced by the attraction of the near opposite charge on the other plate. The earth connection is absolutely essential, for, if it were omitted, there would be two charges in the lower plate, one on the top and one on the bottom, whose actions on the upper plate would nearly counterbalance each other, so that the "condensing" action could not occur. If too large a charge is put on the condenser, the air will become ionized, a spark will occur, and the condenser will be discharged. Evidently more charge may be put into a condenser the larger it is and the closer its plates are to each other, so long as no spark occurs. Also, a much larger charge can be retained in a condenser if the air between the plates is replaced by some other medium, such as oil, or glass, through which it is hard to make a spark pass.



(*a*) FIG. 20-13 (*b*)

Dielectric constant. Leyden jar. As already mentioned (p. 304), the force F between two small bodies bearing charges e and e', whose centers are a distance d apart, is $F = ee'/d^2$, if the experiment is imagined to be performed in a perfect vacuum. Under more ordinary conditions, $F = \frac{1}{K} \times \frac{ee'}{d^2}$, where K is a quantity (due to Faraday) characteristic of the medium between the bodies, now known as the ***dielectric constant,*** these non-conducting media being called "dielectrics." For air $K = 1.0005$, which is so nearly unity as to justify the use of the formula for experiments in air with K omitted. But, if electrostatic forces are measured in oil, through glass or other substances, we find values of K, as given in Table XXV, some of which are quite large. If the forces are thus reduced, the electric field intensity is reduced, and the likelihood of a "breakdown of the medium" becomes much less.

Thus, if a condenser is to be made which will hold a large charge safely, its plates may be immersed in oil of a high dielectric constant; or they may take the form of tin-foil sheets covering oppo-

TABLE XXV

Dielectric Constants

Glass	5–8
Mica	5.8
Oiled paper	2.0
Paraffin	2.1
Quartz	4.4
Water	81
Alcohol	26

site sides of a glass plate, or the inside and outside of a bottle. In the latter form we have a "Leyden Jar" (Fig. 20–14), which may have a central knob and rod connecting with the inner tin foil by a chain resting on the bottom, thus making easy connections possible with both coatings. The outer is the one which is usually connected with the earth.

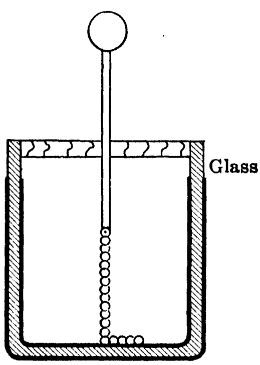


Fig. 20–14
A Leyden jar

The "Breakdown" of a dielectric. Electrostriction. Piezo-electricity. If the air "breaks down," we know that the cause is due to ionization. In the case of a Leyden jar, if excessive charges are put on the coatings, the glass will be punctured by a spark, and the jar thereby reduced to a practically useless air condenser. Even with moderate charges, the glass is under a mechanical strain, whose existence can be shown by means of polarized light (p. 579). The nature of this strain is not hard to imagine, if one considers the electrical nature of atoms and molecules. The positively charged particles will be pulled in one direction, the negative ones in the opposite. It may be whole atoms which are thus strained apart, or parts of atoms. The resulting changes slightly alter the dimensions of the body, a phenomenon which goes by the name of ***electrostriction.*** In the case of certain crystals it is

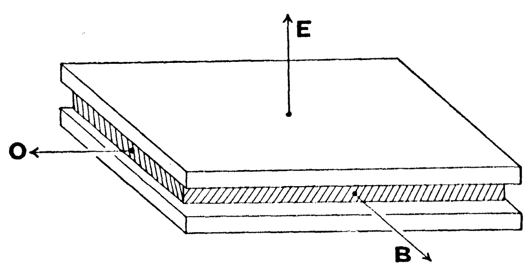


Fig. 20–15
A crystal plate mounted for a demonstration of the piezo-electric effect

quite considerable in amount. Quartz, for instance, if cut in the form of a plate in a particular direction with respect to its crystal faces, and mounted between two metal plates (as shown in Fig. 20–15) will expand along the line E and contract along the line B when the plates are charged like a condenser. The line O must be the direction of the optic axis (p. 577). If the charge on the plates oscillates, the quartz plate can be made to oscillate also, the necessary condition being that the exciting alternations of charge must agree with the natural period of the crystal, as in other cases of resonance. This effect is one of the class known as ***piezo-electric*** ("piezo" refers to pressure). It may be reversed; in which case one puts a strong pressure on the crystal along the line B and charges appear on the two faces of the plate.

To return to the breakdown of a dielectric, one has only to suppose that the strain in the atoms of the material becomes so great that some of the particles are torn loose and that these then move with such violence as to detach a mass of atoms, thus making a hole through the material. The glass (if we take this material as example) has normally no free electrons and is thus a non-conductor, but it might be fair to say that it becomes ionized under the action of the electric field, and thus a current flows through it during the very brief time while the ions are free.

Capacity. A condenser may be thought of as holding electricity, somewhat as a pail holds water. If the plates of the condenser are close together, it corresponds to a wide tub and is said to have a large capacity; if the plates are far apart, it is like a tall narrow pipe. If we had a pail made of some plastic material we could change it gradually from a low wide shape to a tall narrow one, and if it had water in it, this would be shallow in the tub, but deep when the vessel took its tall narrow form. The pressure of the water on the bottom would increase greatly during this same change. In the same way the "electrical pressure" (which is the potential difference) in a condenser which has a certain charge in it is low when the plates are close (large capacity) and high when the plates are widely separated (small capacity).

The water analogy is not quite satisfactory. If one tries to put more water into a pail than it holds, the excess harmlessly spills over. To make it behave like an air condenser, we should have to imagine the pail so arranged that the moment it became full, it tipped over and emptied itself completely. More and more

charge can be put into an air condenser until the potential rises high enough to ionize the air between the plates, when a spark occurs and all the charge is lost. To make a pail resemble a glass condenser, we should have to imagine the bottom to fall out at the moment when the pail became full. In the case of the Leyden jar the glass is punctured, and the jar ruined.

Thus a condenser in ordinary use is never "full" of electricity and the capacity must be defined without reference to such an idea. The larger the charge, Q, which is put into it, the more the difference of potential, V, rises between its plates. We define the ***capacity*** of the condenser (sometimes called capacitance) as the ***ratio of the quantity of charge to the difference of potential produced by it;*** thus

$$C = \frac{Q}{V}.$$

Hence if a large charge produces a small difference of potential the capacity is large, and ***vice versa.*** A condenser has unit capacity (one E–S unit) when a unit charge creates a unit difference of potential between its plates; or, we might say, it creates unit potential on one plate if the other is connected to the earth and thus kept at zero potential. When the condenser plates are close together, the binding of the charges by their mutual attraction is high and much more charge can be put into it before its potential rises to a certain value. Thus it has a higher capacity. In fact it can be shown that in the case of a plate condenser

$$C = \frac{KA}{4\pi d}$$

where A is the area of one of the plates, d their distance apart and K the dielectric constant of the medium between them ($K = 1$ for air, approximately).

The ***intensity of the electric field*** inside a condenser is sometimes wanted, as for instance in the oil-drop experiment. Since 4π lines come from each unit of charge (p. 305) Q units, spread over an area A, give $4\pi Q/A$ lines coming from each square centimeter of one condenser plate, and in this case going straight over to the plate opposite, so long as the plates are large compared with their separation. But the intensity of the field is the number of lines per square centimeter; thus throughout this space

$$\text{the intensity} = 4\pi\frac{Q}{A} = 4\pi\frac{VC}{A} = \frac{VK}{d} = \frac{V}{d} \text{ in air}$$

by the formula above. The potential difference V must be measured in E–S units.

Condensers are made in various forms and sizes in practice. In telephone condensers, thin oiled paper is used as a dielectric, and a large capacity can be obtained in a small volume. Where the insulating medium has to withstand high potentials, oil is often used between the plates. Variable condensers are commonly used in radio sets; their plates alternate without touching and the area of the adjacent surfaces can be changed.

Displacement currents in condensers. When a Leyden jar is charged, the strain in the glass which is thereby produced is caused by a slight displacement of positively charged particles in one direction, and negative ones in the opposite. During the extremely short time while this motion is going on, we may say there is a current of electricity due to this displacement of charged particles, which is appropriately called a ***displacement current.*** If the electric field is now reversed, the displacement current occurs for an instant in the opposite direction. If the electric field is continually reversing, as in a rapidly alternating field, the displacement current does the same, with the result that a rapid oscillation appears to pass through a condenser without difficulty. Nothing does pass through, actually, except an inductive action, but the dielectric particles share to a certain extent in the oscillation which the electrons in the conducting parts of the circuit possess; and hence the oscillations occur inside the condenser, as well as on both sides of it.

Hydraulic analogue of condenser action. If a displacement of the particles of the dielectric occurs, it will annul itself when it gets a chance, that is, when the condenser is discharged. The condenser acts in this respect like an elastic body under stress, which yields, but resists, and will come back to its normal condition when it can. An experiment with water can be imagined which offers a close analogy to the condenser. If AB (Fig. 20–16) is a pipe line full of water, and C a chamber divided into two parts by an elastic diaphragm D, no water can flow through from A to B, but a displacement current can occur, due to the yielding of the diaphragm when pressure is put on A. Its elasticity will make it spring back when it can, sending water out through A, whence it came. If a rapidly alternating current of water should surge back and forth through A, the yielding of the diaphragm will pass it on to the pipe B, as though the

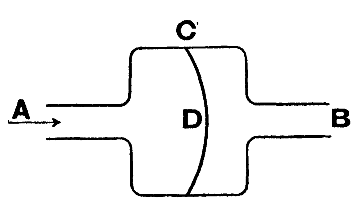


FIG. 20–16

diaphragm were not there at all. Although no water actually flows through the diaphragm the alternating current may be said to pass through it.

Thus many features of condenser action are imitated by this hydraulic model.

Energy of a charged condenser. To put a considerable charge on a condenser we must bring more and more little charges up to it in succession. After the first, these will be repelled by the charge already there, so that work must be done in forcing more charge on the plates. During the whole of this charging process the average potential may be regarded as half of the value, V, finally attained. The work done in carrying a charge Q through an average difference of potential $\frac{1}{2}V$ is $\frac{1}{2}QV$ (p. 317), and this work must be equal to the energy which is thereby stored in the condenser. This will be expressed in ergs, if Q and V are in E–S units. This energy will reappear in the form of heat, etc., if the condenser is discharged.

Condensers in series and parallel. It often happens that condensers are to be used in groups, and the capacity of the groups ought to be known. There are two ways of connecting them, Fig. 20–17*a* in series and Fig. 20–17*b* in parallel.

(*a*) In the ***series*** case the same displacement current flows through all, and it involves the same movement of elec-

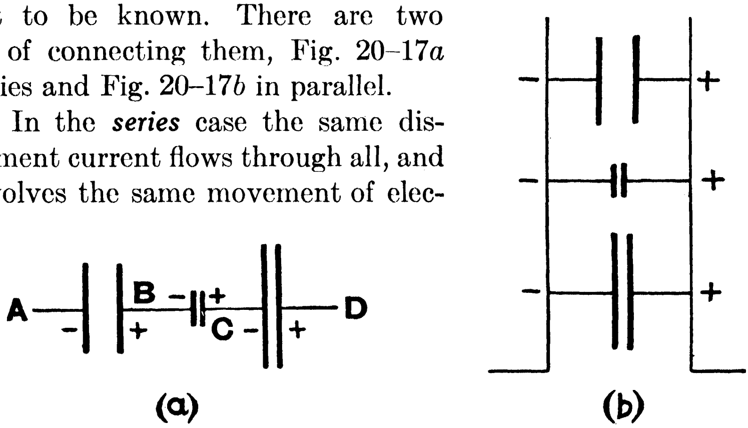


FIG. 20–17
Condensers in groups

trons in each condenser, or in all of them together. The potential, however, rises at each condenser like a series of steps. Thus if the total potential between A and D is V, and the separate condensers whose capacities are C_1, C_2, etc., are charged so that potential differences V_1, V_2, etc. arise between their plates, we have

$$V = V_1 + V_2 + \cdot \cdot \cdot \cdot .$$

In each condenser (or in all together) there is a quantity Q; hence

$$Q = VC = V_1C_1 = V_2C_2 = \cdot \cdot \cdot ,$$

C being the total capacity of the series combination. We may write this

$$\frac{1}{C} = \frac{V}{Q} = \frac{V_1 + V_2 + \cdots}{Q} = \frac{V_1}{Q} + \frac{V_2}{Q} + \cdots = \frac{1}{C_1} + \frac{1}{C_2} + \cdots.$$

Thus the reciprocal of the capacity of the group is equal to the sum of the reciprocals of the individual capacities. (This result is approximate but not strictly exact.)

(*b*) In the ***parallel*** arrangement, the potential difference must be the same in each condenser, since the similarly situated plates are connected together by a conducting wire; but the sum of the charges constitutes the whole charge. Hence

$$V = V_1 = V_2 = \cdots, \text{ and } Q = Q_1 + Q_2 + \cdots.$$

Also $\quad Q = VC,\ Q_1 = VC_1,\ Q_2 = VC_2.$

Hence $\quad VC = VC_1 + VC_2 + \cdots$

or $\quad C = C_1 + C_2 + \cdots;$

that is, the total capacity is found simply by adding the individual capacities of the separate condensers.

Atmospheric electricity. Electrical effects out-of-doors are often distressingly prominent. *Lightning* is an electrical spark on a very grand scale produced by enormous charges borne by clouds. The discharge occurs between masses of clouds or between a cloud and the ground. These charges arise in part, at least, from the motion of air past drops of water, or from water drops breaking up in strong gusts of wind. If the finer drops thus formed are borne aloft by strong ascending currents, they are always found to be negatively charged. The larger drops remain lower down and are positive; thus the upper regions of a "thunder-head" type of cloud are negative and the lower positive, and very large potential differences (up to 100,000,000 volts) may occur if the air currents are sufficiently violent.

The *northern lights* (aurora borealis) are due to electrical effects in the upper air, (50–100 miles up) caused apparently by some action coming from the sun.

The *charge on the earth's surface* is known to be negative. It is possible to study this, and charges borne by the air, by the use of a "collector" placed in the open air and surrounded by ions (by a flame, for instance). The collecting wire will then be electrically connected with the air and will be able to come to the same potential as the region in which it is. When this potential is measured at different heights, it is found to change on the average by 100 volts for every meter of ascent. At times this rate is five times greater, at times much less. Rarely its direction is reversed. The potential is positive (usually) in the air, as compared with the earth. The rate of change of potential (called the poten-

tial gradient) itself diminishes with height, being only one-tenth as rapid at a height of 4 km. (2.5 miles). There is a downward current of positive ions and an upward flow of negative ones, producing a large current (1500 amperes for the whole earth) flowing continually from the air into the earth itself. At any one place the number of these ions is very small and thus the current through each square centimeter is difficult to detect. At present there is no satisfactory explanation of these curious phenomena.

Electrometer. An electroscope may be turned into an electrometer, if a graduated scale can be placed next to the deflected leaf. Such an instrument is relatively insensitive, but if the movements of the leaf are magnified by means of a strong microscope, it may become very useful. If the electroscope is tilted so that the leaf when charged hangs almost vertically, most of its weight then ceases to act against the electrical force producing the deflection, and the instrument becomes more responsive. By further refinements very sensitive leaf electrometers have been produced.

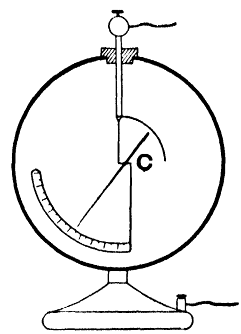


FIG. 20–18

Another convenient form is the Braun electrometer (Fig. 20–18), made with a mechanically balanced metal pointer, pivoted at C, which acts like the leaf of an electroscope, but plays over a graduated scale. These instruments are made in several ranges.

There are several other forms of electrometers, some remarkable for their extreme sensitiveness, others for their quickness of response or their small capacities.

PROBLEMS

1. Two small, equal, conducting balls are charged with $+4$ and -4 E–S units respectively, and are placed 10 cm. apart. Find the intensity of the electric field, in direction and amount, at a point which is 10 cm. from each ball.

2. Two small, equal, conducting balls hung on insulating threads are charged with $+5$ E–S units and -17 E–S units of charge respectively. They are then allowed to touch each other and afterwards they separate to a distance of 4 cm., between their centers. Find the force between them in this position.

3. A and B are two small bodies bearing charges of $+20$ and -10 units respectively, and they are 12 cm. apart. Find the direction and intensity of the electric field at a point C, 10 cm. distant from both A and B. (Graphical solution.)

4. A and B are two small, equal, conducting bodies, bearing charges of $+50$ and -30 E–S units respectively. Find the force between them (a) when they are 10 cm. apart in air, (b) when they are 10 cm. apart in oil of dielectric constant 2, (c) when they are in air, but after they have been touched together and then placed 10 cm. apart.

5. Find the sign and magnitude of the charge on a small body, if its weight appears to be increased by 0.02 gram when it is placed 10 cm. vertically over another small body whose charge is + 100 E–S units.

6. The four corners of a square of 10 cm. side are each occupied by small charged bodies. Going around the square these charges are + 50, − 20, + 50 and − 20 units. Find the resultant force in direction and magnitude which is exerted on one of the charges of − 20 units by the others.

7. Two small, equal, conducting bodies 10 cm. apart are charged with − 6 and + 10 E–S units respectively. A third body charged with + 3 units is placed on a line drawn perpendicular to the line joining the first two bodies at a point midway between them, and 10 cm. out from that point. Find, approximately, the force on the 3-unit body.

8. Two condensers, whose capacities are 10 and 5 E–S units respectively, are connected in series with each other, and the combination is connected to a source of potential difference of 100 E–S units. Find the charge and P. D. for each condenser.

9. Two plate condensers, of capacities 5 and 20 E–S units respectively, are connected in parallel and then charged to a potential difference of 10 units. Find the charge in each condenser, and in the combination, and the total capacity.

10. What is the charge in E–S units which is necessary to bring the plates of a condenser to a difference of potential of 20 E–S units, if the condenser has a capacity of 10 E–S units? If this capacity were then changed to a value 10 times smaller than before, the quantity of charge remaining unchanged, what would the potential become? Suggest a way in which such changes in capacity are produced.

11. Three condensers, whose capacities are 100, 200, and 300 units respectively, are connected together in parallel, and then charged to a potential difference of 40 E–S units. Find the charge in each condenser, and the total charge.

12. Three condensers in series are charged to a total difference of potential of 300 E–S units. If their capacities are 50, 100, and 300 E–S units respectively, find the difference of potential between the plates of each, the charge in each, and the total charge.

13. A Leyden jar resting on a table is to be charged. Its middle knob is connected to one pole of an electrical machine the other pole of which is grounded to a water pipe. Consider how highly charged it would become if (*a*) it rests on a piece of clean glass on the table, (*b*) it is in direct contact with the table surface, (*c*) as a student's hand grasps it, while it rests on the table, (*d*) its outer coating is connected by a wire directly to a water pipe.

14. Will a charged rod attract or repel an ordinary compass needle? Why? Does it make any difference what charge the rod bears, or to which end of the compass needle it is brought near?

15. If in the oil-drop experiment a pair of condenser plates 1 cm. apart were used, and when charged to a P.D. of 6,000 volts they just supported a drop whose weight was 1.1×10^{-10} gram, how much charge must there have been on the drop? (Answer, 11 electrons).

16. Find the capacity of a Leyden jar (essentially a plate condenser) whose conducting plates have an area of 8000 cm.^{2} (each) and are separated by 6 mm. of glass of dielectric constant 6.

17. If the condenser in the last problem is charged to a P.D. of 12,000 volts, how much energy is stored in it?

18. A piece of fur is wrapped around an uncharged hard-rubber rod, and the two together are put inside a metal can which is connected to an electroscope. The rod may be charged by moving it about inside the can, without withdrawing it. Will this create a charge on the electroscope? What inference can be drawn about the equality of the charges produced by friction? If the rod is then withdrawn from the can, leaving the fur inside, what charges will there be, and where?

CHAPTER 21

ELECTRIC CURRENTS

Sources of electricity. Cells. We have already seen that rubbing two different materials together produces electric charges on both, which may be multiplied by such induction devices as the electrophorus. There are many other "sources of electricity." Wires of different materials joined together as in a thermocouple (p. 367) give feeble "thermoelectric" effects when one of the junctions is warmed. Certain crystals show charges when they are compressed (p. 319), or even when they are warmed. Modern commercial methods of "generating electricity" in quantity depend on the induction of currents and are treated in Chapter 25. A common method of obtaining electrical effects (as in flashlights, for instance) is by means of ***voltaic cells,*** named in honor of Volta.[1]

Contact between two different conducting bodies is sufficient to create a difference of potential between them, which is called a "contact difference of potential." Measurements on this rather

[1] A. Volta, (1745–1827), Italian physicist, professor in Pavia. He combined his cells in series in the "voltaic pile," or "crown of cups," and obtained electrical effects of unheard-of violence (for that time); so that Napoleon invited him to repeat his experiments in Paris, and he received many honors. He also invented the electrophorus, and performed the experiment proving that the terminals of a cell are themselves charged. The volt was named after him by an international congress of electrical experts in 1881.

uncertain quantity are made very difficult by changes in its value produced by very thin films of foreign material on the surfaces, which cannot be seen and can hardly be avoided. A layer one atom thick involves very little material and yet it seems to be enough to alter such effects profoundly.

A sure and easy method of obtaining a potential difference between two pieces of different metals is to dip them into a solution, such as a weak acid. This was the essence of Volta's discovery, and we now arrange the experiment as in Fig. 21–1, placing a plate of zinc and a plate of copper (for instance) into dilute hydrochloric acid (HCl) solution. This makes a zinc-copper cell (chemical symbols Zn and Cu). A much better, though more complicated, cell to use is the common "dry" cell, whose plates are of carbon and zinc. The internal action of such cells is considered below (p. 352).

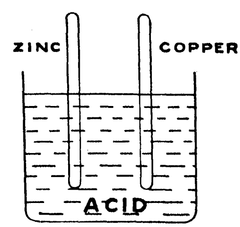


FIG. 21–1
A one-fluid cell

Electrostatic charge on cell terminals. If the terminals of a dry cell are examined in any of the usual ways, no electrostatic charge can be detected on them. If either of them is connected to the leaf of an electroscope, the latter remains unmoved because the potential to which it is raised is too small. An ingenious multiplying device makes it possible, nevertheless, to prove that the plates of a cell are electrostatically charged, and to distinguish the positive from the negative.

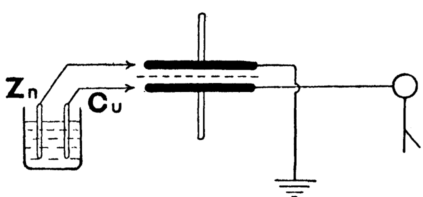


FIG. 21–2
Experiment showing electrostatic charges in a cell

A pair of condenser plates (Fig. 21–2) with good flat surfaces are fitted with insulating handles. The upper one is connected to the earth, the lower to an electroscope, the outer case of which should be grounded. (It is best to have the lower plate a part of the electroscope, but it must not be small.) The plates are separated either by a thin layer of lacquer on the plates themselves (not always trustworthy) or by a thin sheet of paraffined paper, or mica. The upper one may then be connected temporarily with the zinc plate of a voltaic cell and the other at the same time with the copper

plate. The lower plate instantly becomes charged with the same sign as the copper, at a potential different from that of the upper plate by the amount of the P.D. of the cell. So far there is no visible effect on the electroscope. The cell connections having been detached, the upper plate may be lifted off, whereupon the leaves of the electroscope will diverge by an amount easily visible.

The capacity of the condenser is large when the plates are very close together. If we then call it C, we may call it c (meaning a much smaller capacity) when the plates are widely separated. Since a definite amount of electricity Q is placed in it when contact is made with the cell terminals, and this remains unchanged, we may write $Q = vC = Vc$, meaning by this that the potential v is at first very small, but must increase (in order to keep the product equal to the constant charge Q) when the capacity becomes reduced by separating the plates. This increase may easily be made to be of the order of a thousandfold, thus producing a potential high enough to move the leaf of an ordinary electroscope.

When the deflection of the leaf of the electroscope is obtained, it is an easy matter to identify the sign of the charge borne by it. For instance, in Fig. 21–2, the copper plate has been connected to the electroscope and the latter will then be found to be positively charged. A negatively charged hard rubber rod brought down toward it from above will make the leaf collapse by driving electrons into it. If the zinc plate had been the one to which it was connected, the charge would have been negative.

Voltaic cells. Voltaic cells may be made of a great variety of substances and of solutions. The results vary greatly with the solutions used. Some of the most successful combinations are given in Table XXVI.

Magnetic effects of currents. Oersted's [1] experiment. When a copper wire is connected from one terminal of a cell to the other and a compass needle is placed just under one part of this wire, there is a magnetic effect upon the needle which persists as long as the connection is maintained and the cell continues to operate properly. We describe what is going on by saying that a current flows from positive to negative through the wire and that this current produces a magnetic field in the neighborhood of the wire.

[1] First performed in 1820 by H. C. Oersted (1777–1851), professor at Copenhagen.

TABLE XXVI

Name	Negative terminal	First liquid	Positive terminal	Second liquid	E.M.F.
Two-fluid cells:					
Bunsen	Zinc	Dilute sulphuric acid	Carbon	Nitric acid	1.9
Daniell	Zinc	Zinc sulphate	Copper	Copper sulphate	1.08
Grove	Zinc	Sulphuric acid	Platinum	Nitric acid	1.9
One-fluid cells:					
Leclanché	Zinc	Sal ammoniac	Carbon		1.46
Edison-Lalande	Zinc	Caustic potash	Copper		0.98

Actually we now believe that electrons flow from the negative end to the positive, but these terms were adopted at a time when no one could tell which flowed, or whether perhaps both did. It would be better if we could go back and begin over again, calling what moves positive, but no one dares face the resulting confusion. This experiment gives us a first connection between magnetism and electricity; there are many others. No simple explanation can be given of this fundamental experiment. The generation of magnetic fields by currents or by moving charges (p. 337) must be accepted as one of the basic properties of moving electricity.

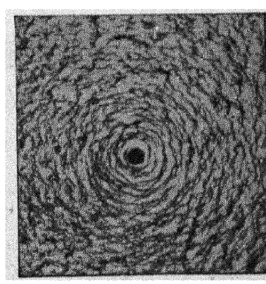


FIG. 21–3
Magnetic field around a current-bearing wire which is perpendicular to the diagram

Magnetic field around a wire carrying a current. If a vertical wire is passed through a small hole in a horizontal plate of glass, and a current is sent through the wire, the glass may be dusted over with iron filings and then the arrangement of the magnetic lines around the current will be disclosed (Fig. 21–3). These are found to take a novel form; they occur as circles about the wire, the stronger part of the field being near the wire, as one might expect. Nothing like this is found in the case of ordinary magnets. It means that an isolated pole, if such a thing were obtainable, would revolve continually in a circle about the wire, an experiment which may be imitated with a real magnet suitably mounted.

If we adopt the usual, but unfortunate, convention that the "direction of the current" is the way positive electricity flows, i.e., from positive to negative, we find that ***a north pole moves around the wire in the same sense as the handle of a screwdriver would if it were driving the point of a screw in the direction of the current.*** This rule gives the relation between the directions of the magnetic lines and of the current, and it will frequently be found useful. Thus, if the point P indicates a wire passing perpendicularly through this page carrying a current going inward, the direction of the lines of force is indicated by the arrows in Fig. 21–4.

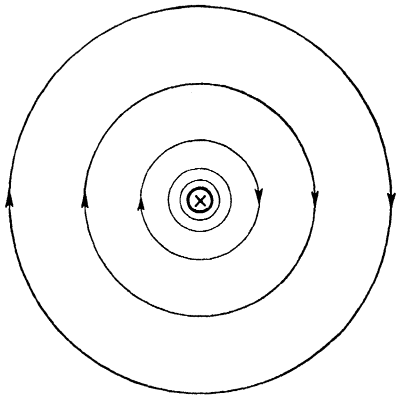


FIG. 21–4

Magnetic field around a coil. If a current flows around a circular loop of wire, the magnetic lines take the form shown in Fig. 21–5*a*. The current in the loop may be supposed to be approaching the observer in the upper wire and going away from him in the lower. The force at points in the plane of the loop is thus perpendicular to that plane, the lines approaching the plane from the left, just as though the loop were replaced by a magnetic sheet, magnetized perpendicularly to its surface with polarity as shown in the figure (*b*). An isolated north pole would then follow curves similar to those shown in (*a*) or (*b*), if it were free to move.

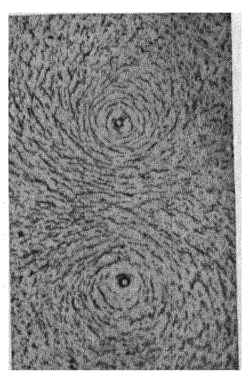


FIG. 21–5*a*
Section of coil

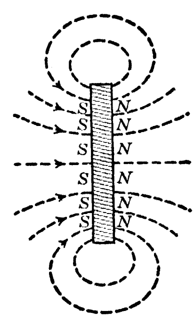


FIG. 21–5*b*

In the case of a coil in the shape of a helix (Fig. 21–6) with a current flowing through it, the magnetic lines take a form almost exactly like those around a bar magnet of the same shape, and if such a coil is freely suspended, it will turn like a compass needle

and be attracted or repelled by magnets or by other coils carrying currents like itself.

The *polarity* of such coils may be found by a simple rule obtained again from the action of a screwdriver. Turning once more to Fig. 21–5, we may imagine ourselves looking at the coil from its left side. The current will then be circulating *clockwise* around the coil, and the magnetic lines going away from us *into the coil.* Thus we see that *if a screwdriver is turned in the direction of the current in a coil, the point of the screw moves in the direction of the magnetic lines.* In the present example the lines are going in toward the face of the coil; that is, an isolated north pole would move that way. Hence the face of the coil in which the current is circulating *clockwise* must be equivalent to the *south* end of a magnet.

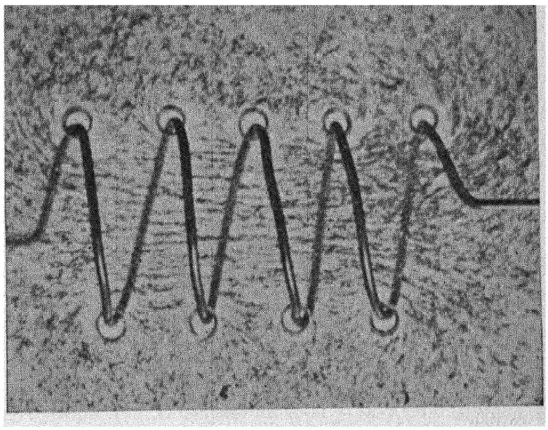


FIG. 21–6

Electromagnetic ("E-M") units of current and charge. The ampere. Ordinary current electricity has so little to do with charges that the unit which it is most natural to use is not that of charge, but is a new one, a unit of current. The magnetic effects of a current are the ones chosen from which to form a definition of this unit. We have no single word used as a unit of current in the case of the flow of water, because there is nothing about water currents analogous to the magnetic effect of an electric current.

The new unit is derived by considering a simple experiment. If a wire loop as in Fig. 21–5a is made with one-centimeter radius and an isolated unit north pole is placed at its center, a current through the loop will exert a force on the pole. If this force is one dyne for each centimeter of wire, or 2π dynes in all for a single turn of wire, the current has then the value known as the *electromagnetic unit of current.* This unit has no universally accepted name; some call it the "ab-ampere" (absolute ampere); we shall refer to it simply as the E–M unit of current. Practically, its chief

importance arises from the fact that from it was derived the universally used unit of current, the ***ampere***,[1] defined by the relation 1 E–M unit = 10 amperes. The definition just given is based on an experiment which is simple in idea but difficult to perform with accuracy. In practice when a current is to be measured, it is necessary to have some easier way of doing it. Thus another method altogether is used from which to derive a workable definition of the ampere, which we shall consider below (p. 365).

The unity of quantity of electricity, or charge, in the ***practical*** system of units is the amount carried by an ampere in one second and is called the ***coulomb***. In the E–M system the unit of charge is the amount carried by 1 E–M unit of current in one second. The following relations exist among these units: 1 E–M unit charge = 10 coulombs = 2.998×10^{10} E–S units of charge.

Calculation of magnetic field from wires and coils. The magnetic field in a circular loop of 1 centimeter radius has 2π units of intensity when 1 E–M unit of current flows through it, as follows from the definition in the previous paragraph. It is not difficult to show that the field produced by a current in a *flat circular coil* is proportional to the current and inversely proportional to the radius of the coil. If H is the intensity of this field, the exact relation among these quantities is

$$H = \frac{2\pi NI}{r},$$

where I is the strength of the current in E–M units and N is the number of turns in the coil, or

$$H = \frac{2\pi Ni}{10r}$$

if i is expressed in amperes. The proof of this formula is omitted.

In the case of a *long narrow helical coil*, whose length is great in proportion to its width, it can be shown that the field inside the coil is given by the formula

$$H = \frac{4\pi ni}{10}$$

where i is the current in amperes, and n is now the number of turns *per centimeter* in the coil.

In each of these formulæ we meet with the product of the current by the number of turns. The term "ampere-turns" is often used as an abbreviated expression for this product. It is an important quantity in determining the strength of the field produced by the coil.

---

[1] In honor of A. M. Ampère (1775–1836), French mathematician and physicist, professor in the polytechnic school in Paris; a man of diversified talents, whose most important scientific work lay in the field of the interrelations of magnetism and electricity.

In addition to flat coils and long thin ones, another case sometimes of use is that of a long, straight wire carrying a current of i amperes. It is an easy calculus problem to show that the intensity of the magnetic field at a distance r from the wire is given by the equation

$$H = \frac{2i}{10r}.$$

The *work done* in carrying a unit pole around in a circle about this wire is (force $\times$ distance)

$$\frac{2i}{10r} \times 2\pi r = \frac{4\pi i}{10} \text{ ergs.}$$

Instruments for measuring currents. Galvanometers. In magnetic experiments such as those described above a current in a coil causes a magnet to turn. Newton's laws of force show us that this action must be accompanied by an equal and opposite reaction. Hence there must be a tendency for the magnet to cause the coil to turn; that is, a torque acts on it. Either part, the coil or the magnet, may be so mounted as to be free to turn in response to this mutual action, and this turning may be made to indicate or measure the current in the coil. Thus we are led to the design of two types of instruments, both called galvanometers, after Galvani,[1] known as the ***moving-magnet*** and ***moving-coil*** types.

The first type is now rare. It can be illustrated well enough for our present purposes by an ordinary magnetic compass set near the center of a large vertical coil. This makes a rough and inaccurate instrument, though by reducing the dimensions of the coil and the magnet and suspending the latter on a thin quartz fiber, a very sensitive and useful galvanometer may be obtained.

The second, or moving-coil type, is common, and is usually known as the D'Arsonval type of galvanometer. In this form (Fig. 21–7) the magnet is of a horseshoe shape with a circular gap in which the very light coil, delicately suspended for the most accurate work or pivoted in the common portable types of instrument, is free to turn. A very small current through the coil creates enough magnetic field to react with the field of the permanent magnet and turn the coil. If the coil is mounted on pivots, as in the figure, two fine coiled springs bring it back to its zero posi-

[1] L. Galvani (1737–1798), physician in Bologna. He discovered that a frog's leg can be made to twitch if its nerve is connected through two wires of different metals to another part of its body. This was supposed to be due to "animal electricity" until Volta discovered its true cause.

tion and also serve to lead the current in and out. If the coil is suspended, the torsion of the supporting wire limits the angle at which the coil stands, and brings it back to rest when the current ceases. The current enters by this same wire and leaves by a loosely coiled one below. Such instruments may be made of extreme sensitivity and are comparatively little disturbed by outside magnetic fields.

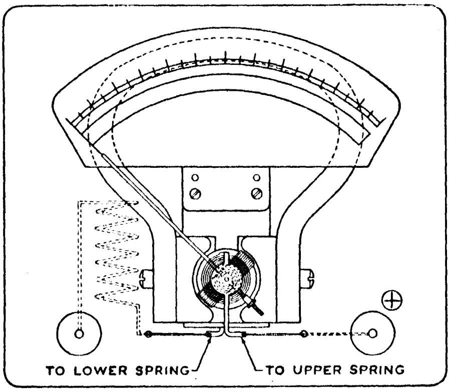


FIG. 21–7
A Weston voltmeter of the D'Arsonval type

Any sort of galvanometer may be transformed with little trouble from a mere indicator of currents into an ***ammeter,*** an instrument for measuring currents, or a ***voltmeter*** for measuring differences of potential (p. 347).

Magnetic effect of a moving charge. It is natural to regard a current as the motion of something. In the case of a "current of electricity" the very term shows that we are thinking of electric charge as a flowing liquid. One way of testing this idea is to try moving a charge and look for the magnetic effect which should accompany it, if it is really equivalent to a current. Faraday states on the basis of reasoning from his lines of force that "if a ball be electrified positively in the middle of a room, and then be moved in any direction, effects will be produced as if a current in the same direction had existed." Rowland [1] was the first to try the experiment, in 1876. He coated a non-conducting disc (Fig. 21–8) with gold foil and rotated it rapidly in its own plane while it was acting as one plate of a charged condenser, the other plate being close to

[1] H. A. Rowland (1848–1901), professor of physics at Johns Hopkins University; a genius of extraordinary mathematical and experimental skill who began his career by carrying out experiments of great value in his own bedroom with home-made apparatus. Helmholtz admitted him to his crowded laboratory as soon as he heard of the important nature of the experiment he proposed to make there (the one described here). Rowland's best-known field of work was his invention and manufacture of concave diffraction gratings of very high quality and the study with them of the spectra of all the elements, and of the sun. He also carried out fundamental researches in magnetism, electricity and heat.

it, and at rest. He found a feeble magnetic effect upon a needle supported at the center of rotation, the rotating charge acting like a current in a coil in the same plane. It made no difference whether the gold foil was in the form of sectors or covered the whole insulating plate. Evidently when the conductor moves, the charge moves with it and acts, as Faraday supposed, like a current in a wire.

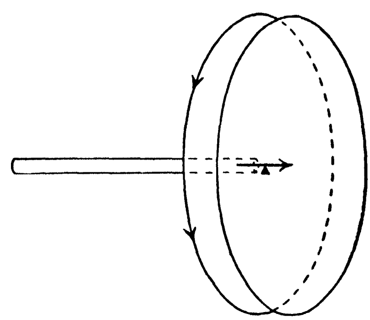


FIG. 21–8
Simple sketch illustrating the principle of Rowland's experiment on the magnetic effect of electric convection

Current through a conductor. Resistance. Ohm's law. If a difference of potential is maintained between the ends of a wire by a cell, there will be a steady current of electricity through the wire, just as there must be a flow of water in a pipe if there is more pressure at one end of it than at the other. The amount of the current, I, will be proportional to the difference of potential E; and it will depend also on the size of the wire. A very narrow wire, like a narrow pipe, offers a poor opportunity for the flow. The current varies, too, with the nature of the wire, copper offering an easy path for it as compared with iron, even of the same size and length. The quality of a particular wire which limits the flow is called its ***resistance,*** R, which is defined by the relation $R = E/I$ or $I = E/R$, and this equation is known as ***Ohm's*** [1] ***law.*** The essence of this statement is that in most cases the ratio of E to I proves to be constant, and hence the resistance is a quantity worth talking about. We shall see that in electron tubes (p. 452) this ratio varies, and the term "resistance" then loses its simplicity. The unit of resistance is called the ***ohm,*** and may be defined as that which permits a current of one ampere to flow through it when there is a difference of potential of one volt between its ends. In practice this definition is made more concrete by saying that the ohm is the resistance of a column of mercury 106.3 cm. long, 1 mm.^{2} in cross-section, at 0° C.

Ohm's law not only serves to define resistance, but it tells us, in the form $E = IR$, that the fall of potential through a conductor

[1] G. S. Ohm (1787–1854), German physicist whose fame rests mainly on this one discovery.

carrying a constant current I is proportional to its resistance. Thus if current is flowing through several wires, one after the other (in "series"), the same current must be flowing through each, and the fall of potential through the set must be in steps whose heights are proportional to the separate resistances. If the wires in which the resistance exists are connected by means of very good conductors, there will be no appreciable potential drop in the connections, so that a diagram showing the fall of potential will resemble a series of slopes with level stretches between, like terraces.

Specific resistance. If wires of different lengths are compared, it is evident that the resistance must have a value proportional to the length. If wires of different diameter are compared, the result is similar to what one would find with pipes carrying water. The ease with which water flows through a pipe depends on the opening it presents to the flow, and this varies with its cross-sectional area, which is proportional to the square of the diameter. Electrons appear to flow through the body of a wire, as water flows through a pipe, and not along the surface only. The ease with which they flow depends on the ***conductivity,*** which is the reciprocal of the resistance.

If wires of the same dimensions but different materials are compared, it is found that the material makes a great difference. All the factors upon which the resistance of a wire depends are combined in one expression

$$R = \sigma \frac{l}{A}$$

where l is the length of the wire, A its cross-sectional area, and σ ("sigma") the value of the resistance when l and A are each equal to unity, i.e., σ is the resistance of a "wire" 1 cm. long and of 1 sq. cm. area (a 1-cm. cube) of the material; σ is called the specific resistance, or the resistivity, of the material. Engineers in English-speaking countries commonly use as a basis for specific resistance not a 1-cm. cube but one foot of wire one "circular mil" in cross-sectional area (i.e., of diameter 0.001 inch). The specific resistance in ohms per mil-foot can be obtained by multiplying the numbers here given by 6,000,000.

Table XXVII shows the specific resistances of a number of metals, and other materials. As resistance is a useful quality in many electrical arrangements, a series of alloys have been developed

TABLE XXVII

Specific Resistances of Elements, Alloys, Insulators

	Resistance of a "wire" 1 cm. long and 1 sq. cm. in area at 18° C.		Temperature Coefficient
Aluminum	3.0	$\times 10^{-6}$ ohms	0.0038
Bismuth	119	$\times$ "	0.004
Cobalt	9.7	$\times$ "	0.0033
Copper	1.7	$\times$ "	0.0042
Iron	9 to 15	$\times$ "	0.006
Mercury	96	$\times$ "	0.00088
Silver	1.65	$\times$ "	0.0040
Tungsten	5.0	$\times$ "	0.0051
German Silver	16 to 40	$\times$ "	0.0004
Manganin (Cu, Mn, Ni)	43	$\times$ "	0.00001
Carbon	0.004 ohm		− 0.0005
Mica	9	$\times 10^{15}$ ohms	
Paraffin	3	$\times 10^{18}$ "	
Sulphur	4	$\times 10^{15}$ "	

which possess it to a high degree. Some of these, like nichrome (an alloy of nickel and chromium), will stand a high temperature in the open air without deterioration, and are used, for example, in electric toasters; some like manganin are useful because their resistance does not vary with the temperature. Table XXVII shows by comparison the specific resistance of some of the common insulators or "non-conductors" and proves that they are at least very reluctant to conduct electricity even if they do not absolutely refuse to do so.

Variations of resistance. Resistance thermometers. The resistance of metals varies with the treatment they have received, and depends slightly in the same material (e.g., steel or copper) on whether it is hard or soft. A more interesting variation, however, is that which occurs with change of temperature. Roughly speaking, the metallic elements (but not alloys) increase in their resistance in proportion to the absolute temperature, and the coefficient of change is of the same order of magnitude as that expressing the expansion of a gas. The gas coefficient is 1/273 or 0.00366 per degree C.; the temperature coefficient of resistance for pure metals varies from 0.003 to 0.006. This odd fact indicates that the electrons in a conductor may bear some resemblance to the gas particles in a tube, but this indication must not be taken

too seriously as the coefficients vary both among conducting wires and expanding gases.

For a given material, the law of change of resistance with the temperature can be accurately determined, and the resistance of a little coil set in the bottom of a tube, like the bulb of a thermometer, makes an accurate instrument for measuring temperatures. This ***resistance thermometer*** (Fig. 21–9), if made of platinum or some other high-melting metal, has almost as wide a range as the gas thermometer, and is far more convenient. Its indications may be read, and even recorded, at some distance from the resistance coil, as long connecting wires offer no special difficulty. Thus, for instance, a series of reading or recording instruments in a central office enables one operator in a large manufacturing plant to watch over a large number of furnaces. This advantage is shared by thermoelectric thermometers also (p. 367).

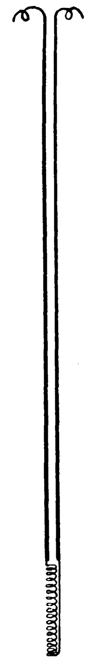


FIG. 21–9

A resistance thermometer. The coil at the bottom is of fine wire, and the whole is encased in a protecting tube.

A fine wire, through which enough current flows to make it quite hot, changes considerably in resistance when a current of air blows across it. This has been applied to measuring the velocity of the wind, and also to the reception of sound signals, as in sound ranging (p. 247).

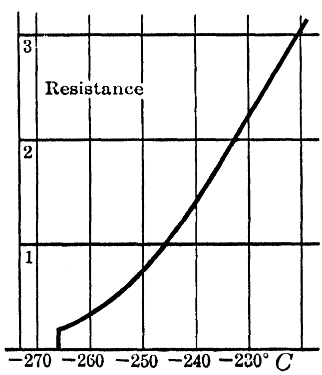


FIG. 21–10

Resistance of lead at very low temperatures in millionths of an ohm per cm.³ At − 266° C., it suddenly falls to zero. At higher temperatures it follows a straight line, rising as though from about − 260° C.

Super-conductivity. When the specific resistances of metals were measured at very low temperatures by Onnes (p. 205), it was found that many of them continued to diminish in a regular way to as low a temperature as could be reached. In the case of certain metals, however, there was at some definite temperature a sudden drop in resistance to practically zero (Fig. 21–10), so that a current which was started flowing in a coil or loop of such a metal continued for a day or two before becoming imperceptible. Lead changes to this "super-conductive" state at 7.2° above the absolute zero, mercury at 4.2°, tin at 3.8° and thallium at 2.3°. Currents can

be started in such coils by the approach or retirement of a magnet (as on p. 386) and can be detected by their magnetic effect. No simple explanation of this remarkable phenomenon has yet been found.

Theories of electrical conductivity. A conductor has already been said to be a body containing free electrons. Many theories have been devised on this basis to explain the different resistances of different substances, and the way in which these change with temperature, pressure and other conditions; but no one of these theories is as yet accepted as completely satisfactory.

It is safe to say that each individual free electron may travel a distance many times greater than the width of a single atom before coming to rest. Its place may then be taken by some other electron, which makes a similar journey, and so on, thus accomplishing the carrying of the electric current. The number of free electrons available need not be more than one for each atom, and only in extreme cases would they all be needed at any one instant.

An interesting hypothesis is that the electrons move mainly in channels among the atoms, along which their motion is comparatively free. The kinetic agitation of adjoining atoms increases with the temperature, and hinders the journeys of the electrons, thus giving rise to the observed increase of resistance. The atoms in crystals are always regularly arranged in space (p. 122); thus free channels might be expected among them. If a metal consists of many fine crystals, it might not be easy for electrons to pass from one to another. Such a metal should have higher resistance than if the crystals were larger; and this is often the case.

The atoms of non-conductors might well lose some electrons if sufficiently agitated by heat, which would then become free, as they are in metals. It is a fact that glass and other insulators become rather good conductors when red-hot. Probably the number of free electrons always varies with the temperature, though very little in the case of metals.

Electrical circuits. An electrical circuit contains one or more sources of potential difference, and resistances in various possible arrangements, together perhaps with measuring instruments. The source of electricity (a cell, for instance) creates a tendency to flow, as a pump does in a water circuit. Electrons flow about through the whole circuit, driven through it by the source, as the water would be in a circulatory system. Often there are alternative paths in parts of the circuit. Figure 21–11 shows such a case as a sample circuit, drawn in the conventional manner. *C* shows the common symbol for the cell, two different plates, the short

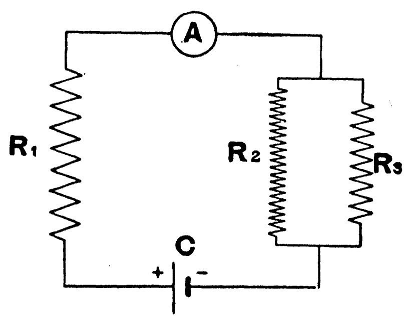


FIG. 21–11
An electric circuit, showing the conventional symbols

thick one being usually the negative terminal. R_1 is a resistance in series, that is, all the current has to flow through it; it is drawn as a sharp zigzag line. A is the symbol for an instrument for measuring the current, an ***ammeter.*** R_2 and R_3 are two resistances in parallel, R_2 being of finer wire and higher resistance than R_3; the current divides at P, part going one way and part the other, reuniting at Q. Naturally the larger part goes by the easier path, through R_3.

The straight lines in the diagram signify connecting wires of negligible resistance.

Combined resistances. It is common to combine resistances ***in series;*** in which case all the current passes through each. The resistance of the combination is then the sum of the separate resistances.

In cases where resistances are arranged ***in parallel,*** as in Fig. 21–11 on the right-hand side, it is easy to find the resistance of the combination. Each path offers a certain conductivity to the current. The conductivity of the combination is the sum of the separate conductivities; just as in pipes carrying water, the total amount carried is the sum of the amounts passing through each pipe separately. Remembering that the conductivity is the reciprocal of the resistance, we obtain $1/R = 1/R_2 + 1/R_3$ where R is the resistance of the combination. A similar rule can be extended to three or more resistances in parallel.

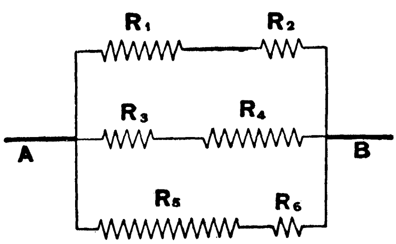


FIG. 21–12

Cases also arise in practice in which ***series and parallel*** arrangements are combined. Figure 21–12 shows one such. To find the resistance between A and B, we add R_1 to R_2, R_3 to R_4, and R_5 to R_6, and then add the reciprocals of these sums, which gives us the reciprocal of the desired resistance. If R is the result

$$\frac{1}{R} = \frac{1}{R_1 + R_2} + \frac{1}{R_3 + R_4} + \frac{1}{R_5 + R_6}.$$

Measurement of resistance. Ohm's law yields a simple method of finding the resistance of a conductor. A steady current is passed through a resistance, R, as in Fig. 21–13. The fall of potential, E, between the points B and C at the ends of R is

measured by means of a voltmeter, Vm (p. 347), connected to them as shown. The voltmeter resistance is so high that a negligibly small current flows through it. The ammeter has a negligible resistance of its own and measures the current, I, passing through the resistance. Then $R = E/I$, which is the value desired.

If the resistance of the voltmeter is not high compared with R, it draws an appreciable current, and it is important that the connections be made as shown, so that the ammeter will record only that current which passes through R. If R is very low, the resistance of the ammeter may not be negligible in comparison, in which case the voltmeter should be connected between B and D, so as to measure the P.D. between the ends of R itself. This method requires a current large enough so that the ammeter can measure it with some precision, but as ammeters can be obtained with different ranges, this is not in itself a serious difficulty. If a large current must be used, however, the heat created by it in the wire introduces changes in the resistance to be measured, as explained below. If the wire being measured is thin, like one of those in an incandescent lamp bulb, for instance, it might even be made red-hot by the current.

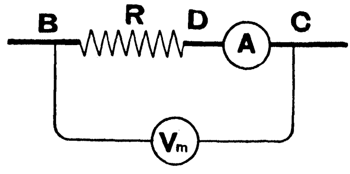


FIG. 21-13
Measurement of resistance by means of voltmeter and ammeter

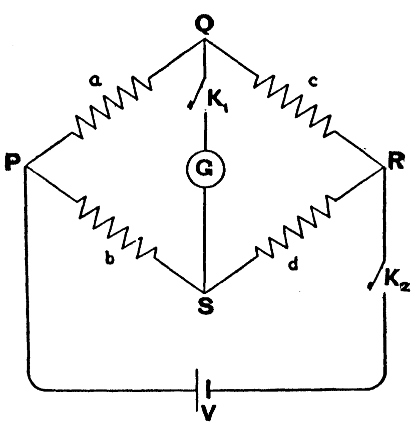


FIG. 21-14
The Wheatstone bridge circuit

A more accurate method of measuring resistances depends on a "bridge circuit," invented by Christie (1833) but usually known as the Wheatstone[1] bridge. In Fig. 21-14, a, b, c, and d are resistances, G is a galvanometer, K_1 and K_2 are keys which stand open until we wish the current to flow; pressing them completes the circuit. V is a voltaic cell. If P is connected to the positive terminal of the cell, it is like a high point on a hill, and going from P to R via Q or S is in either case going down hill. If no current flows through QS when the

[1] In honor of Sir Charles Wheatstone (1802–1875), an inventive genius who was the practical founder of modern telegraphy, and did notable work in many branches of physics.

through a nozzle inside a pipe creates a rush of water past *A*, the base of one tall pipe, driving it over toward another pipe at *B*, so that a difference of level is maintained in the tubes, as at *D* and *E*, which corresponds to the potential difference between the plates of a cell. At *C* is a cock allowing, or preventing the flow through a connecting tube, which stands for the outside circuit in the electrical case. If *C* is shut, the maximum difference in height is attained, corresponding to the electromotive force of the cell. If *C* is slightly opened, a small current flows through the "outside circuit" and the jet below is no longer able to maintain quite the same difference of level; in the electrical case the potential difference between the plates of the cell falls a little, and is no longer equal to the electromotive force. If a large opening at *C* is made, the difference of level may fall very much; if the connecting tube is very wide and freely open, the jet mechanism is incapable of maintaining *any* considerable difference of level. This corresponds to the electrical case of a "short-circuit" between the plates of the cell, when the potential difference practically disappears.

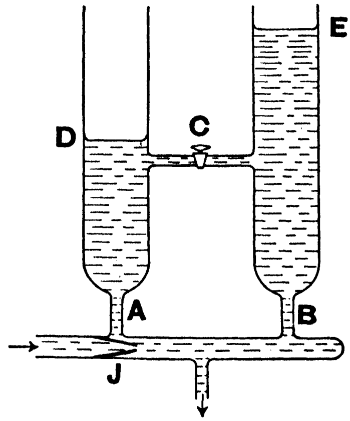


FIG. 21–15
The hydraulic analogue of a cell

Ammeters. Any galvanometer may be made into an ammeter or into a voltmeter. The essential points of each instrument must now be considered.

An ***ammeter*** must measure a current without offering any obstruction to its flow, that is, without having any considerable resistance. As the moving coil of the usual well-made galvanometer is made of fine wire, it cannot avoid having some resistance, too much, in fact, for this purpose. To solve this difficulty a "shunt" or coarse wire of very small resistance is connected between the terminals of the coil, as shown schematically in Fig. 21–16. *A* and *B* are the terminals of the instrument to which the wires of the circuit are fastened. Between them, inside the case of the instrument, is the shunt connection, and also the connection to the moving coil. Thus there is a pair of parallel resistances between *A* and *B*. If *S* is the resistance of the shunt

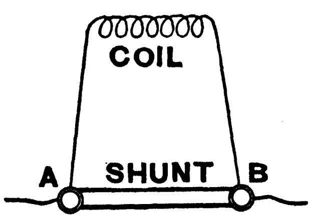


FIG. 21–16
The circuit inside an ammeter

and C that of the coil, the resistance of the combination is given by R, where

$$\frac{1}{R} = \frac{1}{C} + \frac{1}{S} \quad \text{or} \quad R = \frac{CS}{C+S} = \frac{S}{1+\dfrac{S}{C}}.$$

The last expression shows that if S is small compared with C the resistance of the combination is nearly that of the shunt resistance itself, that is, it is very small. Small as it is, however, there is still a small difference of potential between A and B when a current flows, for by Ohm's law ($E = IR$) this must always be the case if there is ***any*** resistance at all. This difference of potential will be proportional to the current, and it will drive a small current through the moving coil. This coil may easily be made so sensitive that a very small current will turn it through a large angle. If a pointer is attached to the moving coil so as to move over a graduated scale, the reading on the scale will be proportional first to the current in the moving coil, and thus to the difference in potential between A and B, and finally to the current in the main circuit. Hence the scale may be graduated to indicate the main current, even though the measurement is actually made by means of a feeble current in a side circuit.

Shunt circuits. In the last paragraph a shunt circuit was shown to be useful. It is often desired to reduce a current by 10, 100, or 1000 times, for convenience of measurement. If, for instance, a current is flowing through a part of a circuit containing a resistance R and it is required that this shall be reduced to 1/100th of its value, one has only to connect a shunt r as shown in Fig. 21–17, and the current will divide between the two branches in inverse proportion to the resistances, the larger current flowing through the smaller resistance. If the whole current in the main circuit be considered as divided into 100 parts and one of these is to be sent through the branch containing R, there will be 99 left to go through r. The current must thus divide into two parts whose ratio is 1 : 99, and thus r/R must be equal to 1/99. Similarly if 1/10 of the current is to be made to go through R, r/R must be 1/9.

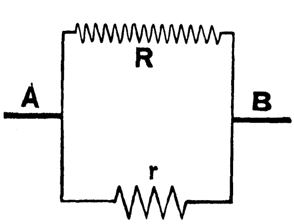


FIG. 21–17
A shunt

Voltmeters. A voltmeter is made for the purpose of measuring a difference of potential, and must do so without destroying, or even altering it. The hydraulic analogy of page 345 shows that if much current is drawn from a reservoir, the height of the reservoir is altered. Similarly a voltmeter must not draw much current from the source of E.M.F.; it must therefore have high resistance, and be so delicately made that the very feeble current passing through it is able to move its coil adequately. Sometimes the coil has a high enough resistance already, but more often extra resistance must be put in, so as to bring the total up to a proper value. Figure 21–18 shows the outer terminals *A* and *B* of the voltmeter, the moving coil and the added resistance. To change the scale of a voltmeter and make still higher "voltages" (differences of potential) readable on it, we may add still more resistance.

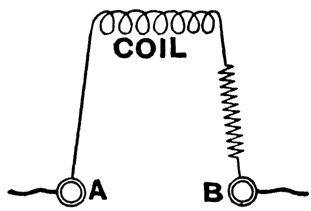


FIG. 21–18
A voltmeter

For example, if the coil of the instrument has a resistance of 100 ohms, one volt put on it will produce a current of 0.01 ampere. If the instrument is so sensitive that 1 scale division deflection of the pointer is produced by 0.0001 ampere, then we shall have 100 divisions deflection for 1 volt. If we wish only 1 division per volt, we must increase the resistance to 100 times as much. i.e., we must add 9900 ohms in series with the coil.

Applications of Ohm's law. Ohm's law may be applied to an entire circuit, or to any part of it. A couple of examples will make this clear.

Example 1. If an electrical generator creates a difference of potential of 120 volts between the two wires of a (direct-current) circuit at the power station, how much difference of potential will there be at a building a short distance away if the transmission line has a resistance of 0.02 ohm, and the various electrical devices in the building are using (*a*) a current of 200 amperes, (*b*) no current?

The fall of potential between the power house and the building is equal to $e = IR$, where R is the resistance between these points, and I the current flowing through this resistance. Hence the fall in case (*a*) is $0.02 \times 200 = 4$ volts, and the difference of potential at the building is thus 116 volts. In case (*b*) $I = 0$, and there is no fall of potential; thus the full potential difference, 120 volts, exists at the building.

Example 2. A circuit is arranged as Fig. 21–11, p. 341. The cell is supposed to be a dry cell, furnishing an E.M.F. of 1.5 volts. It has no resistance of its

own worth mentioning. $R_1 = 7.6$ ohms, $R_2 = 6$ ohms, $R_3 = 4$ ohms. Find the current indicated by the ammeter, the difference of potential between the ends of the pair of parallel resistances, and the current through the 4-ohm coil, R_3.

The whole current cannot be found unless the whole resistance is known. The resistance of R_2 and R_3 combined is given by

$$\frac{1}{R} = \frac{1}{6} + \frac{1}{4} = \frac{5}{12}; \quad R = 2.4 \text{ ohms.}$$

The total resistance in the circuit is $7.6 + 2.4 = 10$ ohms. The total current

$$I = \frac{E}{R} = 0.15 \text{ ampere.}$$

Applying Ohm's law next to the part of the circuit including R_2 and R_3, the fall of potential e through this group is given by $e = IR$ where e, R and I refer to the same part of the circuit, i.e., $e = 2.4 \times 0.15 = 0.36$ volt. We may then apply Ohm's law again to the branch through the 4-ohm resistance; thus,

$$\text{current through the 4 ohms} = \frac{\text{potential difference between its ends}}{\text{resistance of that branch}}$$
$$= \frac{0.36}{4} = 0.09 \text{ ampere;}$$

which completes the problem.

Of course, the rest of the current, i.e., 0.06 ampere, must go via the 6-ohm resistance, and the two currents are in the ratio of 6 to 9, which is the inverse ratio of the resistances, as it should be.

The potentiometer. An important electrical circuit is the one used in the ***potentiometer,*** an instrument for the accurate measurement of electromotive forces. Its arrangement is indicated in Fig. 21–19. A battery of cells, B, furnishes a steady current through a resistance RT composed of resistance boxes or slide wires or both. A side circuit makes contact with RT at two movable points, P and Q. This side circuit contains a sensitive galvanometer, G, and either one of two cells C or W, depending on how the switch arm, S, is connected. C is the cell whose E. M. F. is to be tested; W is a standard cell of known E. M. F. When S is connected to C, it is possible to place the points P and Q so

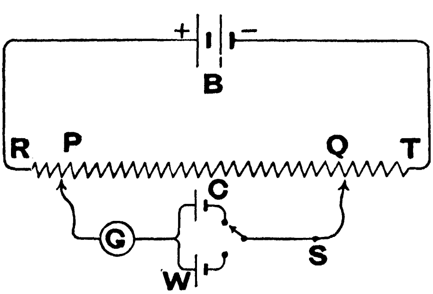


FIG. 21–19

that the fall of potential between them due to the current in RT is exactly the same as that produced by the cell C in the side circuit. Hence no current will flow through the galvanometer or the cell C at all. If r_1 is the resistance in PQ under these conditions and r_2 the resistance similarly obtained when the cell W is connected instead of C, the E.M.F. E_C of the cell C is given in terms of E_W, that of the standard cell by the relation

$$\frac{E_C}{E_W} = \frac{r_1}{r_2},$$

since the resistances are proportional to the differences of potential. It is essential that the current drawn from B should remain constant throughout.

The potentiometer may also be applied to many other sorts of measurements. It is to be noted that this method draws no current from the cell at the moment when it is being tested. One can, therefore, be sure that no such effect as polarization (p. 353) is coming in to make the results inaccurate; and the precision of the results may be made extremely high by choosing a very sensitive galvanometer.

Networks. A more complex circuit in which *steady* currents are flowing can be solved by means of two common-sense statements, known as Kirchhoff's[1] laws. These are:

1. At any point in a network the sum of the currents entering and leaving must be equal to zero. (Of course, no accumulation of electricity can occur at any point.)

2. Tracing out any closed loop in a network the potential must rise in one part and fall equally in the rest so as to return to its original value at the starting point. If no source of E.M.F. is present in this loop, the sum of the positive IR potential changes must be equal to the sum of the negative ones. If there are sources of E.M.F. in the loop, their sum must be equal and opposite to that of all the IR potential changes added together.

PROBLEMS

1. In the experiment (p. 329) demonstrating the existence of an electrostatic charge on the plates of a voltaic cell, both plates are usually fitted with insulating handles. What difference would it make if the upper plate had a metal handle?

2. Find the specific resistance of a metal, given that a wire 2 meters long and of 1 mm. diameter has a resistance of 1 ohm.

3. Find the resistance of a column of mercury at 0° C. which has a length of 106.3 cm. and a cross-sectional area of 1 sq. mm.

---

[1] See footnote, p. 532.

4. Find the diameter (in mm.) which a manganin wire must have in order to possess a resistance of 1 ohm per meter.

5. Find the resistance of 1 mile (1610 m.) of trolley wire (copper) which is $\frac{3}{8}$ in. (9.5 mm.) thick.

6. If we have three coils whose resistances are 2, 3, and 4 ohms respectively, we can make 16 different resistances from these by using them singly or in groups in different ways. Find the value of each of these resistances.

7. Find the resistance of a coil if there is a difference of potential of 16 volts at its ends when a current of 2 amperes is flowing through it. What would be the difference of potential if the current were reduced to half an ampere?

8. If a trolley wire has a resistance of 0.4 ohm per mile, find the drop in potential over a 2-mile length of it when a current of 25 amperes is flowing through it.

9. It is desired to reduce the current through an instrument whose resistance is 25 ohms to one-fiftieth of its present value. How can this be done by means of a shunt, assuming that the current in the rest of the circuit is unaffected by this change?

10. A current of 26 amperes is flowing through a circuit, part of which consists of two resistance coils in parallel, of 1 and 12 ohms resistance respectively. Find the current in each coil.

11. A galvanometer of 100 ohms resistance is to be used as a voltmeter, but it is much too sensitive. It is proposed to shunt it with a wire, so as to reduce to one-hundredth the amount of current going through the instrument itself. Find the resistance of the necessary shunt wire, and indicate whether or not such a scheme would work out satisfactorily.

12. Show how to change a galvanometer, which has a resistance of 20 ohms, and gives a deflection of 1 division for 0.001 ampere, (*a*) into an ammeter reading 1 ampere for 10 divisions; (*b*) into a voltmeter reading 1 volt per division.

13. A cell, whose E. M. F. is 2 volts, with no resistance of its own, is connected with two resistances A and B in series. One of these is a single 10-ohm coil, but the other, B, consists of 3 coils of 6, 4, and 3 ohms respectively, connected in parallel. Find the difference of potential between the terminals of B and the current through the 3-ohm coil.

14. A closed circuit is formed of a 12-volt storage battery, of negligible internal resistance, and two resistances A and B in series. B consists of two 10-ohm resistances connected in parallel. How large a resistance must A have in order that a current of 1 ampere may flow through it, and what will a voltmeter then read whose terminals are connected to the ends of resistance A?

15. A potentiometer (see Fig. 21–19, p. 348) has a uniform wire RT of resistance 20 ohms and of length 150 cm. The distance PQ is 110 cm. when the standard cell W (of 1.08 volts E. M. F.) is thrown in, and 74 cm. with the unknown cell. Find the E. M. F. of the unknown.

CHAPTER 22

CHEMICAL EFFECTS OF CURRENTS. THERMOELECTRICITY

Electromotive force of cells. The seat of the electromotive force in a cell is usually at the surfaces of contact of the metals with the solution. If a piece of ordinary zinc is dipped into dilute acid, there is an action at the surface of the metal which makes the atoms of the metal come off into the solution, and in so doing they leave electrons behind, thus becoming positively charged. This action produces what is sometimes called a "solution pressure," analogous to the vapor pressure above a liquid. The zinc atoms in the solution having been deprived of electrons (two each in this case) are called positive ions. The zinc metal retains the electrons which the ions leave behind, so that it becomes negatively charged. This charge soon rises to a value high enough to prevent the escape of any more positively charged zinc atoms from the metal.

If a cell is made of zinc and copper rods dipping into an acid solution, the zinc becomes negative with respect to the solution, but the copper acts differently. Positive ions are deposited on it, so that it takes a potential higher than that of the solution; thus the potential rises abruptly from the zinc metal to the solution just outside it, remains constant throughout the solution if no current is being drawn from the cell (or falls somewhat when a current flows) and rises abruptly again at the copper. The internal

actions in such a cell are really very complex and it is impossible to discuss them here. We need only note that the greatest difference of potential between the two terminals of the cell is called the electromotive force of the cell, and is obtained only when the cell is idle. If a current is being drawn from the cell, the difference of potential is always lessened.

Solutions and ionization. In a solution such as dilute hydrochloric acid, HCl,[1] there is an action which makes the compound break up into ions, H positive and Cl negative, so that the solution becomes filled with active charged bodies of both signs, ready to move, and in so doing to form electric currents. Such solutions are conductors of electricity, but the ions are not very free to move, hindered as they are by the overwhelming abundance of water molecules; thus the velocity of the ions is usually quite slow and considerable electrical resistance is involved in this sort of electrical conduction. The reason why the molecules break up into ions in solutions is probably because the electrostatic forces holding the atoms together are weakened. Coulomb's law (p. 318) shows that this force is reduced in substances of high dielectric constant, and this constant for water is extremely high.

Action of a simple voltaic cell. When zinc (Zn) and copper (Cu) are dipped in dilute HCl, the zinc ions come off in the solution near the zinc terminal (or "electrode"), and these repel the positive ions which are already present in the solution, so that the ones near the copper plate go almost at once to it, some of them being deposited on it, charging this plate positively. The forces which keep the hydrogen and chlorine ions apart in the solution (whatever these are) are also responsible for pulling the zinc ions away from the metal and for keeping the chlorine ions from combining with and neutralizing the zinc ions in the solution. Such a condition is at any rate maintained on the whole, however individual ions may be acting.

The remarks so far made apply to the case of a cell when it is inactive. But when an outside circuit is formed, and a current

[1] For the benefit of students who may not have studied chemistry, it is necessary to say that the atoms of different elementary substances are indicated by more or less self-evident symbols as in this case, and their compounds by these symbols set side by side. Here an atom of hydrogen, H, combines with an atom of chlorine, Cl, to make a very active gas, HCl, which is readily soluble in water, forming an acid solution.

flows through it, the situation changes. The accumulated electrons on the zinc terminal have a chance to escape. This removes the attracting force holding the zinc particles from escaping into the solution, so that the zinc metal is gradually used up and must be replaced in time if the cell is to be kept in long-continued action. During all this action the positive ions in the solution (H) are deposited on the copper, and there neutralized by the electrons which have traveled around the outside circuit from the zinc to meet them. They thus gather as bubbles of hydrogen on the copper and so reduce the surface in contact with the solution, or, they substitute a hydrogen surface for a copper surface. This effect is called ***polarization,*** and prevents a plain cell of this description from being very useful, even when it is freshly made.

The dry cell. A description in somewhat similar terms can be made to apply to all forms of cells. The common "dry" cell contains a wet paste including materials for the prevention of polarization; and carbon is used in place of copper for the positive terminal. The zinc forms the outer surface and makes a container for the rest of the cell. When this is worn thin by the action of the cell, the solution escapes and commonly ruins the cell. The two "plates" have large surfaces in order to reduce the internal resistance.

The Weston cell. A cell of particular interest, called the Weston cell, is used as a practical standard of E. M. F., for the reason that it is so easily made, and gives always the same value for the E. M. F., if the materials used are pure. It is made of a glass tube in the form of an H (Fig. 22–1). At the bottom of each leg of the tube is a wire sealed into the glass, serving as an electrode. Inside, enveloping the wire on the negative side, is a cadmium-mercury amalgam; on the other side mercury (Hg) covered with a paste of mercurous sulphate (Hg_2SO_4). Filling the rest of both vertical tubes, and the cross tube, is a saturated solution of cadmium sulphate. This cell gives an E. M. F. of 1.0183 volts at 20° C., which changes extremely little with the temperature. This result furnishes the ***legal definition of the volt*** in the United States (p. 365).

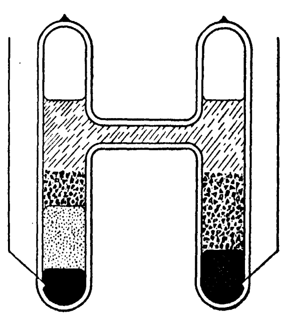


FIG. 22–1
The Weston Cell

Hill diagrams for cells. If one represents difference of potential by difference of height, one may, as already indicated, make a sort of "hill diagram" as a graphical representation of the differences in potential in a cell, as in Fig. 22–2. In this diagram, h is the difference of potential between the two terminals, so long as no current is being drawn from the cell. It is therefore what we have agreed to call the electromotive force (E. M. F.) of the cell.

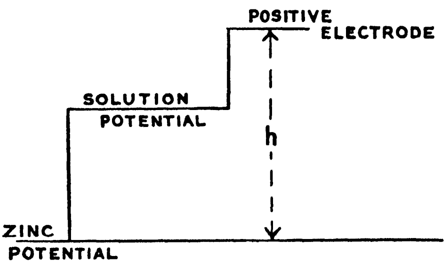


FIG. 22–2
Hill diagram for an inactive cell

When current is being drawn from a simple cell (a dry cell, for instance) the mechanism responsible for the rise of potential at the surface of the zinc is so active and effective that the solution there is as highly positive as before. The hill diagram for such an arrangement takes the form of Fig. 22–3. Starting at the left, the zinc metal is the lowest (most highly negative) point. The solution just outside is positive; the potential falls through the solution, in accordance with Ohm's law, which says that the fall of potential is equal to Ir where r is the internal resistance of the cell, and I is the current. There is, as before, a rise at the surface of the positive electrode, and the height h is the difference of potential between the two electrodes, or the reading of a voltmeter connected to the terminals. (It is, however, no longer the E. M. F. of the cell.) The potential then falls in the outside resistances through this height h, back to the level of the zinc terminal again. The diagram should be imagined to be cut out and the right-hand end bent around so as to meet the left, and complete a circuit.

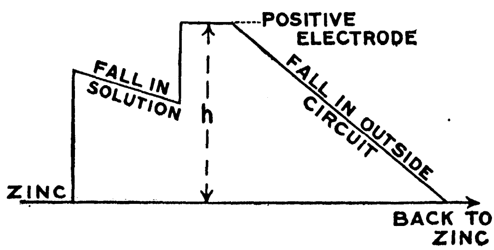


FIG. 22–3
Hill diagram for an active cell

If the cell is ***short-circuited*** by a wire of negligible resistance, the hill diagram takes the form shown in Fig. 22–4*a*. There is now no difference of potential possible between the two terminals, but the

solution near the zinc is as positive as ever. The whole fall of potential occurs inside the cell, due to its internal resistance.

In using the hill diagram in connection with problems, the rise at the positive plate of the cell may be combined with (i.e., added to) the rise outside the negative plate, and not separately indicated in the figures. This can introduce no error in any ordinary case. Figure 22–4*b* shows how this may be done for a short-circuited cell.

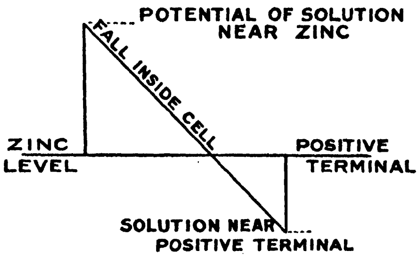


Fig. 22–4*a*

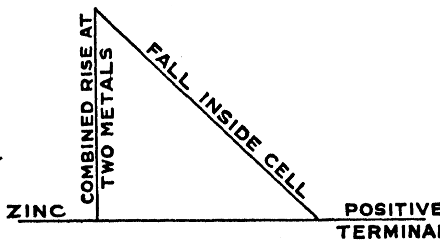


Fig. 22–4*b*

Hill diagram for short-circuited cell

Examples on cells. The action of cells will become clearer by considering a few examples.

Example (1). A dealer tests a dry cell by short-circuiting it through an ammeter of negligible resistance. If the current is 25 amperes, and the E. M. F. 1.5 volts, what is the internal resistance of the cell?

The internal resistance r is all there is to limit the current.

$$I = \frac{\text{E.M.F.}}{r}, \quad \text{or} \quad r = \frac{1.5}{25} = 0.06 \text{ ohm.}$$

If this cell were old, the internal resistance would be higher and the maximum current less.

Example (2). The same cell is connected through a 1-ohm external circuit. Find the current and the reading of a voltmeter connected to the terminals of the cell.

The total resistance is now external plus internal, which adds up to 1.06 ohms. The current is $1.5/1.06 = 1.415$ amperes. The fall of potential through the external circuit $(IR) = 1.415 \times 1 = 1.415$ volts, which is the reading of the voltmeter. In addition, the fall of potential inside the cell is Ir or $1.415 \times 0.06 = 0.085$ volt. The total fall of potential is $0.085 + 1.415 = 1.5 =$ the E. M. F., as it must be.

Storage cells. The types of cells hitherto considered are called "primary" cells, since they furnish current directly at the expense of chemical energy, and can be renewed only by the replacement of the "worn-out" materials. We now turn to the very useful

storage or "secondary" cells, in which the chemical changes can be reversed by forcing current through the cells in the proper direction. In the "charging" process, the chemicals are manufactured anew which were used up in the ordinary running of the cells. To call this process "charging" is a little misleading, as it implies the storage of electrical charges in the cell, instead of chemical energy; but, as chemical reactions are probably almost entirely electrical, it is perhaps superfluous to be too particular about fine distinctions.

The common storage cell is the ***lead cell.*** In this (Fig. 22–5)

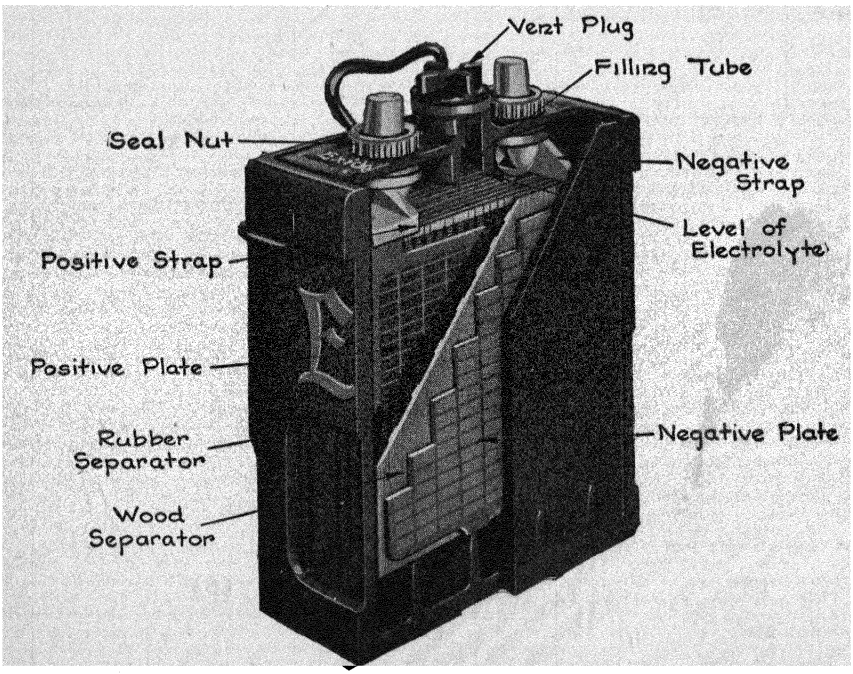


FIG. 22–5

A lead storage cell. (Courtesy of the Electric Storage Battery Company)

two plates dip into dilute sulphuric acid (H_2SO_4). The negative one is of lead and the positive of lead peroxide. The active materials are in a spongy mass, forced into many little cells in a formed lead plate. In use the lead peroxide changes to lead sulphate and the lead changes largely over to lead sulphate also, so that the plates become alike, and the cell is "discharged." "Charging" it turns the negative one back into lead, and restores the peroxide in the positive plate. The porous masses of active material have a large surface, thus greatly increasing the amount of

accessible material. By making the plates large and placing them close together the internal resistance is reduced. The positive plate is brown; the negative gray. When the cell is run down, the sulphuric acid is more dilute (used up in forming the sulphate) and the density of the solution is low. Figure 22-6 shows the changes in density that occur as a cell is discharged. A hydrometer (p. 12) is often used to test the condition of the cell. When fully charged, the density is 1.27 and the E. M. F. about 2.05 volts. As a machine the cell is about 75% efficient; i.e., about one-fourth of the energy used in charging it is lost, largely in heat.

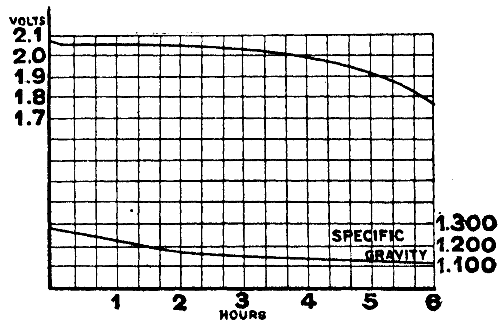


FIG. 22-6
Changes in E. M. F. and in specific gravity of the solution in a lead storage battery

Storage cells are commonly rated in ampere-hour capacity. Thus one of 100 ampere-hours would give a current of 10 amperes for 10 hours or of 2 amperes for 50 hours. In the best modern cells, the maximum current that can be drawn is limited only by the internal resistance of the cell, or by the safe carrying capacity of the external circuit. The E. M. F. of the cell falls with use according to the discharge curve in Fig. 22-6. This fall is more marked with large currents, i.e., a 100-ampere-hour cell would drop farther if discharged at a 100-ampere rate for one hour than at 10 amperes for 10 hours. After many chargings and dischargings the porous material will in time become loose, and fall to the bottom of the cell where it may accumulate to such an extent as to reach to the plates and short-circuit them, unless plenty of room is provided for it.

The Edison cell. The Edison cell is a storage cell of considerably less weight than the lead cell. Its solution is an alkali, caustic potash, KOH, its positive terminal nickel hydrate, $Ni(OH)_2$, and its negative plate iron oxide. Both substances are held in pockets or chambers in nickeled steel plates. Its E. M. F. is 1.3 volts. While its construction is more expensive than that of the lead cell, it stands neglect and abuse better (e.g., impure solutions) and has a larger capacity for the same weight.

Cells in groups. Batteries. A battery is a group of cells. These groups are commonly made up either of dry cells or of storage cells. In either case the internal resistance is likely to be low, and the usual reason for combining them is to obtain a high E. M. F. The best arrangement for this purpose is to put them in ***series.*** In this way, each cell added raises the difference of potential between the extreme plates in the series, which becomes nE, if n is the number of cells and E the E. M. F. of any cell alone. The current obtained in an outside circuit will then be proportional to the number of cells, so long as the internal resistance remains low compared with the external. But the internal resistance when many cells are connected in series may rise to a considerable amount, since it, like the E. M. F., is added when the cells are connected in this way. Hence to cover all possible cases it is safer to reckon in both the external and internal resistances when calculating the current from a series battery.

If the cells are combined in ***parallel,*** the effect is equivalent to increasing the size of the plates of one cell, and thus to reducing the internal resistance without, of course, altering the E. M. F., since the group is now like one large cell. Thus the current for the parallel arrangement becomes

$$I = \frac{E}{R + \frac{r}{n}}$$

by Ohm's law. The parallel arrangement is advantageous only when the internal resistance is considerable in comparison with the external, as may occur when large currents are to be sent through low resistances.

A ***combination*** of both methods of grouping cells is often used. Cells may be put in series to raise the E. M. F., and several such groups may be connected together in parallel for the purpose of lowering the internal resistance. In such arrangements each series group may be treated separately (that is, its E. M. F. and resistance determined) and then these groups can be combined in parallel.

Illustration of grouping of cells. If it is desired, say in a lecture experiment, to produce a current of at least 100 amperes for a short time through a wire of 0.02 ohm resistance, can this be done with ordinary dry cells, and if so, how many are needed and how must they be arranged? Let us assume each cell to

have an E. M. F. of 1.5 volts and an internal resistance of 0.1 ohm. One cell can produce

$$I = \frac{E}{r} = 15 \text{ amperes}$$

through a negligible external resistance. Thus to get 100 amperes would require at least 7 cells in parallel. But the external resistance is not negligible in this case, so that more cells will be required to drive so large a current. Suppose that a group of 10 cells in parallel was tested. Its internal resistance would be one-tenth that of a single cell, and therefore = 0.01 ohm. Its E. M. F. would be 1.5 volts. Four such groups in series would have an internal resistance of 0.04 ohm, and an E. M. F. of 6.0 volts. Through an external resistance of 0.02 they would send a current of

$$I = \frac{6.0}{0.04 + 0.02} = 100 \text{ amperes.}$$

Thus a total number of 40 cells arranged in this way would do the thing required. But if 36 cells were arranged in three groups in series, 12 cells in parallel in each group, the same current would be obtained with fewer cells. We could *never* get the desired result by putting all the cells in series, for even with an infinite number of them, it is easy to see that the maximum current would be only 15 amperes. Nor should we succeed if all the cells were in parallel; for an infinite number of them would then give us 75 amperes. Thus we can solve the difficulty only by combining the two methods of grouping.

If we had storage batteries available, we could do with fewer cells. A 300-ampere-hour lead storage battery, for instance, has an E. M. F. of 2 volts per cell and might have an internal resistance of only 0.01 ohm. We could get 100 amperes through a wire of 0.02 ohm with two of these storage cells in series, and over 170 amperes through the same wire if we had three cells in series and two such groups in parallel, as the reader may easily verify.

Electrolysis. If we set up a cell containing two plates of the *same* metal, dipping into a solution, it will not be a voltaic cell. Any difference of potential that may be developed at one plate will be balanced by an equal and opposite action at the other. If, now, we connect the plates to some outside source of E. M. F., we may force a current through the solution, and whatever positive ions there are in it will move in what we call the direction of the current, while the negative ions will move in the opposite direction. Each set of ions on arrival at one of the plates will deliver its charges and become neutral; i.e., the ions are "deposited" at the plates, according to their nature. If they are gaseous, they form bubbles; if metallic, they form a coating, furnishing an example of electroplating. Whatever is formed, the whole process of passing a current through an ionized solution is called ***electrolysis.*** The

vessel in which this is done is often called an ***electrolytic cell.*** This must not be confused with the voltaic cell. The latter is a source of current, while the electrolytic cell is one in which separation and deposition of different materials out of a solution are accomplished by means of a current.

Electroplating and other applications. In case it is desired to plate a piece of a conducting material with a coating of some metal, the composition of the solution should be such that only one sort of positive ion is present in it; namely, the ion of the metal in question. If, for instance, we wish to deposit a coating of nickel, it is necessary to start with a solution of a nickel salt (usually nickel ammonium sulphate). A nickel rod or plate is chosen for the positive terminal (Fig. 22–7). Then, as nickel ions are deposited out of the solution at the negative terminal, others come away from the nickel terminal, and keep the solution uniform. The negative plate, on which the deposit is to be laid is called the ***cathode;*** the positive one is the ***anode.***

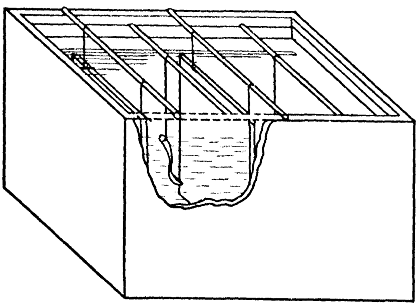


FIG. 22–7

An electroplating bath. The current enters by bars laid across the bath, on which are hung plates of the metal. Opposite to these are the objects to be plated, hung on other bars insulated from the first set.

The rate at which the metal is deposited is important; if it comes on very slowly the crystals may be too coarse; if too fast they may not adhere well. Usually the ion concentration in the solution should be very low. There are great variations in the proper conditions with the different metals.

Another common application of electrolysis is in making copper plates from which books are printed. The type with which each page is first set up serves only to make an impression in wax. This is coated with graphite to make it conducting, and a thin sheet of copper is then deposited all over this surface. After the wax is removed, the copper sheet is backed by a strong layer of type metal and this forms a durable "electrotype" plate with which the printing is actually done. Seamless tubing also is made by a similar electrolytic method.

The electrolytic process of refining metals is very extensively

used; copper for electrical purposes, for instance, is prepared in this way. By adjusting the strength of the solution, and of the current, copper can be deposited while other metals present as impurities fall to the bottom of the cell. Thus the metal can be obtained in a high state of purity. Aluminum is similarly made, being deposited from a bath of molten ore. Even rubber can be electrolyzed from its fluid "latex," as the particles suspended in this liquid acquire charges, and can be deposited upon a positive electrode.

Faraday's laws of electrolysis. The electrochemical equivalent. In experiments such as those just described it is found that the amount of metal deposited depends on the current, and, of course, on the time also. If several electrolytic cells are connected in series, and the current is depositing copper, for example, in each, the weight of copper deposited is found to be the same in each cell no matter what the size of the plates or the nature of the copper salt used in the solution. This fact was discovered by Faraday along with others of fundamental importance. His results may be summarized thus:

(1) *The same amount of a given sort of metal is always deposited by the same current in the same time; or by a smaller current in a longer time.* More exactly, when a given amount of metal is deposited, a *definite quantity of electricity* is associated with this operation; from which it follows that *each atom must carry a definite charge with it.*

(2) When the same current flows through cells in series in which *different* metals are being deposited at the same time, the weight deposited in each is characteristic of the metal in that cell.

The *electrochemical equivalent* of the metal is defined as the amount of metal deposited when each unit of electricity flows through the cell. The unit chosen for this definition is the *coulomb* (p. 334). If all positive ions were charged to the extent of lacking one electron only (as happens with silver and with hydrogen), the electrochemical equivalents would be proportional to the weights of the atoms. But many metal ions bear two or more of these positive charges, that is, they are atoms deprived of two or more electrons. The ions of copper or mercury may be singly or doubly charged under different chemical conditions; iron ions may

be doubly or trebly charged; tin doubly or quadruply; and so on. The number of charges borne by the atom is the same as the number to which chemists have given the name of ***valency.*** Thus doubly charged copper ions lack two electrons and are said to have a valency of two. The electrochemical equivalent of copper in this state, that is, the weight of copper deposited when a coulomb of electricity flows through the cell, is evidently half what it is when the copper bears only a single charge. In general the ***electrochemical equivalent*** is proportional to the ***atomic weight divided by the valency.***

TABLE XXVIII

Electrochemical Equivalents

Material	Atomic Weight	Valency	Electrochemical Equivalent in Grams per Coulomb
Hydrogen	1.008	1	0.00001045
Silver	107.88	1	0.0011183
Copper	63.52	1	0.0006589
		2	0.0003295
Tin	119	2	0.0006167
		4	0.0003084
Nickel	58.68	2	0.0003041
Zinc	65.37	2	0.0003387
Oxygen	16	2	0.0000829

Table XXVIII gives the values of the electrochemical equivalents of a few elements; the metals and hydrogen are positively charged as ions, oxygen negatively.

The charge borne by an ion. These facts of electrolysis need be no surprise to anyone accustomed to the ideas of the electron theory of atoms. A positively charged metal ion is simply one formed from a metal atom which normally contains loosely bound electrons, but has been deprived of one or more of them. Each metal has its definite habits in this respect, depending on how firmly its outer electrons are attached to the rest of the atom.

These discoveries of Faraday, though made long before the electron was known, furnish one of the strongest pieces of evidence in its favor. They prove without a doubt that each ion carries its charge in little unit packages, which cannot (under these circumstances, at least) be subdivided. To make a direct connection with the electron theory, we have only to add that recent experiments have shown that the charge borne by a hydrogen or a silver

ion in electrolysis is exactly equal to that borne by an electron, though of opposite sign. One method of finding the ionic charge might be described.

Experimental determination of ionic charge. A current is sent through acidulated water, using platinum strips as terminals or "electrodes" (Fig. 22–8). In this case hydrogen is "deposited" at the negative plate (cathode) and oxygen at the positive plate (anode). The gases can be collected and measured. From what the kinetic theory has already given us we can deduce the actual number of atoms of hydrogen that have been deposited. The product of the current in amperes by the time in seconds during which the current flows gives the number of coulombs of electricity transported by these hydrogen ions. Hence we find the number of hydrogen atoms required to carry one coulomb, and conversely the amount of charge borne by each atom. We find as a result that 1.008 grams of hydrogen (this number is its atomic weight) carry 96,470 coulombs of electricity. This same number of coulombs, known as the Faraday constant, is carried by 107.88 grams of silver, or by a number of grams equal to the atomic weight of any other substance, divided by its valency. There are 6.06×10^{23} atoms in this amount of hydrogen, and hence each atom must carry 1.59×10^{-19} coulombs, or (since 1 coulomb $= 3 \times 10^9$ E–S units) 4.77×10^{-10} E–S units. But this last result is exactly what Millikan in his oil-drop experiment found to be the charge borne by an electron (p. 309). Thus we find very strong support for the electron theory of atomic structure in the phenomena of electrolysis.

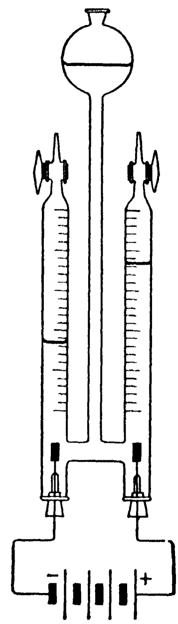


FIG. 22–8

The ions in solutions. The facts of electrolysis are readily explained on the assumption that there are positively and negatively charged ions in the solution. The question arises as to how these come to be formed, and why the atoms do not remain intact and uncharged. It seems likely that the high value of the dielectric constant (p. 319) of water weakens the electric forces binding the atoms together. Whether this is the reason or not, one important conclusion can easily be reached, namely that the process of forma-

tion of the ions is complete before the current is turned on. If the number of ions depended on the applied E. M. F., the current would depend on the E. M. F. once on this account and again as it usually does according to Ohm's law; in other words, it would actually vary as the square of the E. M. F. But Ohm's law holds in solutions as well as in conducting wires. Hence the ions must be formed on account of forces acting in the solution itself. The evidence points to the conclusion that a solution (of common salt, for instance) contains nothing but ions when it is dilute, but that when it is concentrated, there may be a considerable proportion of salt molecules which are not ionized.

These ions move in a solution under an applied electromotive force in opposite directions. Much work has been done on their speeds. Usually they travel very slowly, less than a hundredth of a millimeter per second if driven by a potential difference of 1 volt per centimeter. This slow speed would accomplish very little in the way of electric conduction if it were not for the enormous numbers of ions formed, which may easily be of the order of 10^{20} per cubic centimeter.

Review of electrical units. We have seen that there are three systems of electrical units in existence, two of them based on the centimeter-gram-second (C. G. S.) system. These two are the electrostatic (E–S) and the electromagnetic (E–M) systems. The third is a practical compromise between the first two. It will be a convenience to have summaries of these three systems for purposes of reference and ready comparison. The E–S system is little used except for purely electrostatic calculations. The E–M system is mainly of interest in theoretical work, and because the practical system was derived from it. The fundamental unit in the E–S system is that of charge, while in the other two it is current. In each case the other units are derived from the fundamental one.

TABLE XXIX

Electrical Units

The Electrostatic (E–S) Units

Charge	Two unit charges on small bodies whose centers are 1 cm. apart act on each other with a force of 1 dyne.
Current	In the E–S system we rarely deal with a flow of current, but we could make a unit by supposing unit charge to pass a point in 1 second.

Difference of potential	One erg of work is involved in transferring unit charge from one point to another if their difference of potential is unity.
Capacity	The unit is a capacity of such size that it is raised to unit difference of potential by unit charge.

The Electromagnetic (E–M) Units

Charge	The unit, derived from that of current, is the charge carried by unit current in one second past a given point.
Current	The unit current flowing in a circular loop of 1 cm. radius, made of one turn of wire, produces 2π units of magnetic field intensity at the center of the loop, or, it exerts a force of 2π dynes on a unit pole placed there.
Difference of potential	The unit is defined in the same words as in the E–S system; i.e., it is such that one erg of work is involved in transferring unit (E–M) charge through it.
Capacity	The unit is a capacity of such size that it is raised to unit difference of potential by unit charge.

The Practical System of Electrical Units

Charge	The *coulomb* is the charge past a given point carried by one ampere in one second. It is equal to the charge borne by 6.3×10^{18} electrons.
Current	The *ampere* is one-tenth of the E–M unit of current. For convenience it is usually measured by the silver deposited by electrolysis. The "international ampere" is the current which will deposit 0.0011180 gram of silver per second. Since this value has been adopted, it has been found that the value which is proper for one-tenth of the E–M unit of current is 0.0011183. The difference is usually unimportant.
Difference of potential	The *volt* is nearly 1/300 (better 1/299.8) of the E–S unit, or 10^8 times the E–M unit. Strictly speaking, this definition gives the "absolute volt," which differs from the international volt by a small quantity (one part in 30,000) which is usually negligible. The international volt (the one in ordinary use) is defined by the statement that the E. M. F. of a standard Weston cell at 20° C. (p. 353) is 1.0183 volts. The volt was adopted at a time when cells were almost the only source of electricity, and the E. M. F. of the commonest type then in use (the Daniell cell) was nearly one volt.
Capacity	The *Farad* (after Faraday) is the capacity which is raised to a difference of potential of 1 volt when 1 coulomb of charge is put in it. This unit is so large that one millionth of it, the microfarad, is in common use.

Inductance The unit of inductance (p. 393) is the *Henry*. It is the inductance in which a rate of change of current of 1 ampere per second induces an E. M. F. of 1 volt. It also is a large unit.

Resistance The *Ohm* may be defined as the resistance through which an E.M.F. of 1 volt will drive a current of 1 ampere. For practical purposes, the "international ohm" was agreed upon as the resistance at 0° C. of a column of mercury 106.300 cm. long, of 14.4521 grams weight and uniform cross-section (about 1 sq. mm.).

The ratio of the units. The "ratio of the two C. G. S. systems of units" is an important and interesting number. By this term is meant the number 2.998×10^{10} which is in several cases the ratio of a unit of one system to the corresponding unit of the other. For instance, this number is equal to

$$\frac{\text{E–S unit of potential difference}}{\text{E–M unit of potential difference}} = \frac{\text{E–M unit of charge}}{\text{E–S unit of charge}}$$

It is to be noted that sometimes one, sometimes the other system contains the larger unit. This same number is of theoretical interest since Maxwell [1] (1864) showed that it must be equal to the velocity of light in centimeters per second. This remarkable conclusion of the electromagnetic theory must be taken for the present without proof but it has been accurately verified by experiment. It furnishes the first bond between electricity and light.

THERMOELECTRICITY

Electricity from heat. We derive electricity commercially from power, much of which is itself derived from heat. It would be a great saving if we could produce electricity directly from heat and so we can, but only in relatively small and disappointing amounts. The phenomena involved are, nevertheless, both interesting and useful.

If we take two wires, one of copper, the other of iron, and twist or solder their ends together, we form a thermoelectric circuit. If a

[1] J. C. Maxwell (1831–1879), mathematician and physicist; professor in the University of Cambridge. He gave mathematical form to Faraday's ideas on the action of magnetic and electric fields, and developed the electromagnetic theory of light, making possible the discovery of electromagnetic waves. His name is also associated with a color disc, a special form of gyroscopic top, and many other experimental devices. His humorous verses on solemn scientific subjects are worth knowing.

sensitive galvanometer is inserted, say in the middle of the copper wire, Fig. 22–9, we shall find no current flowing if the two copper-iron junctions are at the same temperature; but if one is warmed, or the other cooled, a small current flows through the galvanometer and lasts while the temperature difference is maintained. This fact was discovered by Seebeck in 1822. The direction of the current at ordinary temperatures is such that electrons flow from iron to copper through the hot junction and oppositely through the cold junction. The conventional agreement as to the direction of the current refers to the flow of positive electricity, which would, of course, if it really occurred, be in the direction opposite to that taken by the electrons.

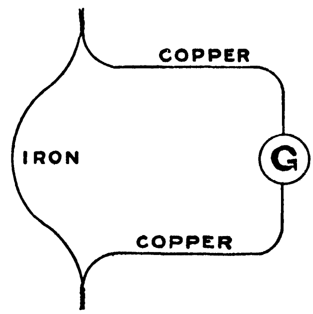


Fig. 22–9
A simple thermoelectric circuit

With the copper-iron "thermocouple" just described the amount of this current increases for a while as the temperature of the hot junction is steadily raised, reaches a maximum, declines to nothing and eventually reverses when the hot junction approaches a red heat. Other combinations of metals give very different effects both in regard to the amount of the thermoelectric E. M. F. and the way it changes in the ordinary range of temperature.

Thermoelectric thermometers. Certain alloys have been discovered, for instance those going by the names of chromel (chromium and nickel) and alumel (aluminum and nickel) which when used as the wires in a "thermocouple" give an E. M. F. almost exactly proportional to the difference of temperature between the two junctions, if that of the cold junction is kept constant. A couple made of such materials, attached to a sufficiently sensitive voltmeter, becomes a ***thermoelectric thermometer*** (or "pyrometer"), and the scale of the voltmeter may be graduated so as to read temperatures directly. The E. M. F. is always small, so that the voltmeter should respond appreciably to a thousandth of a volt in order to be useful. (Such an instrument is usually termed a ***millivoltmeter***.)

Thermometers of this type are simpler to use than resistance thermometers, and are just as convenient and as good, excepting possibly for work of the very highest precision. The E. M. F. of a

thermocouple made from two alloys is sensitive to the slightest variation in composition of either alloy, to the presence of impurities, and even to variations of hardness, etc., due to mechanical or heat treatment; so that great precautions must be taken in order to get two junctions exactly alike. Two pieces cut from the same spool of wire will in most cases not give identical results. Exposure to furnace gases must be avoided on account of the chemical changes produced by them in the wires.

Thermocouples as practical sources of current. If a large number of thermocouples made of substances giving relatively high E. M. F. are put in series, as in Fig. 22–10, the alternate junctions being hot, then cold, then hot and so on, the thermoelectric E. M. F. of each will be added to that of the others, as in the case of cells in series. Thus by using several hundred junctions a considerable E. M. F. may be obtained. In order to draw much current from such a combination, its resistance must be small, that is, its wires must be made thick, heavy, and short, in which case they conduct a great deal of heat from the hot junctions to the cold ones. Everything considered, the resulting battery is inefficient and expensive to run. When fed with a large gas flame, such a generator, made of 120 junctions, has been made to yield an E. M. F. as high as 8 volts and a current, through a negligible external resistance, of 2.5 amperes, but only 0.5% of the heat given to the generator reappeared in the form of available electrical energy.

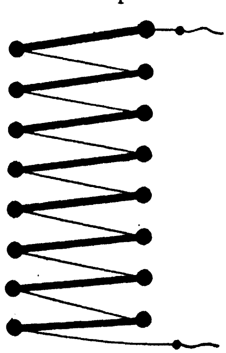


FIG. 22–10

Thermocouples as radiation receivers. The wires of a thermocouple may be made very fine, and the junctions may be hammered out to very thin sheets. If these are then blackened, and exposed to steady heat radiation, they will almost at once attain their final temperature, perhaps a few thousandths of a degree higher than before. By connecting them to a very sensitive galvanometer, changes of temperature of the order of a hundred-millionth of a degree may be measured, and such an instrument becomes a delicate and trustworthy recorder of radiation. Many junctions may be connected in series (Fig. 22–11), with the "hot" junctions all exposed to the radiation, and arranged in close array in a short straight line, while the "cold" junctions are placed on either side, shielded from the radiation. With such an instrument, made of suitable wires, and associated with a concentrating mirror, the heat radiated by a man could be detected at night from a dis-

tance of a mile away. Somewhat similar instruments have made the heat from some of the brighter stars measurable.

Peltier effect. Thomson effect. In 1834 Peltier[1] found that if a current from some outside source is forced through the two junctions of a thermocouple, one is heated and the other is cooled. This cooling has even been used to freeze water. It is easily shown by the following arrangement: if a double-throw switch is so connected that a thermocouple (or better a series of thermocouples, known as a "thermopile") is connected first to a dry cell when the switch handle is thrown in one direction, and then to a delicate galvanometer instead when the handle is quickly reversed, the galvanometer will show for a while a current due to the temperature difference created in the two junctions by the current from the cell.

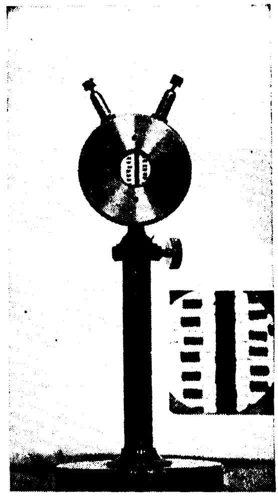


FIG. 22–11

A convenient and sensitive "radiation thermopile." On the right is seen an enlarged picture of the junctions and fine connecting wires.

Other effects connecting heat and electricity have been observed; for instance, a flow of electric current through an unequally heated bar encourages or hinders the flow of heat depending on the direction of the current with respect to the heat flow and the nature of the metal. This is known as the *Thomson* (Lord Kelvin) *effect.*

Thermoelectric effects in ordinary circuits. An ordinary laboratory circuit is likely to involve copper wires, brass terminals and possibly other metals also, such as special resistance alloys, etc. If the entire circuit is at one temperature, no thermoelectric effects can occur; but this almost never happens, especially as one usually warms parts of the apparatus by handling them. Thus it is a frequent experience to find in a circuit small electromotive forces which are not intended to be there. Their effects can often be eliminated, but these details are left for the laboratory.

PROBLEMS

1. If the E.M.F. of a lead cell is 2.1 volts (on open circuit), while the terminal voltage when the cell is delivering 10 amperes is only 2.0 volts, what is the internal resistance of the cell?

2. A lead cell has an E.M.F. of 2.1 volts and an internal resistance of 0.004 ohm.

[1] J. C. A. Peltier (1785–1845), French watchmaker, who devoted the latter part of his life to scientific pursuits.

(a) What will be its terminal voltage when discharging 25 amperes?

(b) What would be the outside applied voltage necessary to charge the cell at the rate of 25 amperes? Draw a hill diagram of this arrangement.

3. A cell of E.M.F. 2 volts and internal resistance 0.5 ohm is connected to the ends of a resistance of 1.5 ohms and at the same time a voltmeter is connected to the terminals of the cell. Find the reading of the voltmeter.

4. Four dry cells, each of 1.5 volts E.M.F. and 0.1 ohm internal resistance, are connected in series, and the group is connected in series with a resistance of 0.1 ohm. What current flows through the resistance, and what would be the reading of a voltmeter connected to the terminals of the battery while the current is flowing?

5. A group of 10 dry cells is connected in series. Each cell has an E.M.F. of 1.5 volts and an internal resistance of 0.1 ohm. This group is then connected in series with two resistances *A* and *B*. *A* consists of a single 7-ohm resistance. *B* is a group of two resistances connected in parallel, of 3 and 6 ohms respectively. Find the reading of a voltmeter, whose terminals are connected (*a*) to the terminals of the battery, (*b*) to the two ends of resistance *B*. (*c*) Find the current through the 3-ohm resistance.

6. A storage battery (6 cells in series) is to be charged on a 110-volt direct-current line, using a charging current of 10 amperes. Each cell has an internal resistance of 0.02 ohm, and an E.M.F. during charging of 2.2 volts opposing the charging current. Show just how much resistance must be placed in the circuit, and how it is to be connected.

7. Six dry cells (E.M.F. of each 1.5 volts; internal resistance 0.06 ohm) are connected in two parallel groups, each group consisting of 3 cells in series. They are used to light three tubes (all alike) in a radio set, the tubes being in parallel with one another. If the resistance of each tube is 6 ohms and the current through each should be 0.0625 ampere, how many ohms resistance should be put in the circuit?

8. A circuit shown in Fig. 22–12 contains two 6-volt storage batteries of negligible internal resistance and two resistances, of 4 and 2 ohms respectively. Draw a hill diagram of this circuit, and find the reading of a voltmeter connected (*a*) between *B* and *C* and (*b*) between *C* and *F*.

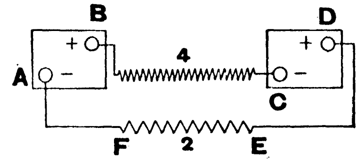


FIG. 22–12

CHAPTER 23

HEATING AND MAGNETIC EFFECTS OF CURRENTS

Power consumed in a circuit. Even to one already familiar with the household uses of electricity it is an interesting experiment to send a considerable current through a long horizontal nichrome wire (the material used in electric toasters), and to see it sag as it expands, then begin to glow visibly, until finally it reaches a bright red incandescence. The amount of heat radiated out by it seems so great to one standing near that it is not surprising that a large fraction of a horse-power is needed to keep it in this condition. Several questions may arise from this simple experiment; for instance, one might inquire what arrangements were best in order to get heat from electric currents, why we do not all heat our houses that way, how best to get light out of such a wire, how to get light without heat, and even why the heat is developed in this way at all. We shall proceed to examine some of these questions.

The cause of the development of heat in a wire is to be ascribed to an action like friction. Everyone knows that if water is being pumped through a pipe, friction uses up some of the energy. In like manner work is required to drive electricity through the resistance which a wire offers to its passage. It is not difficult to calculate this loss from electrostatic ideas.

A difference of potential was defined (p. 317) as being measured by the work (in ergs) done in carrying unit charge through it. If a current I carries a quantity of electricity Q in time t through a difference of potential E, the work done will be equal to QE ergs if these quantities are all measured in E–S units. But the E–S unit of charge is less than the coulomb by a factor of 3×10^9, and the E–S unit of potential difference is greater than the volt by a

factor of approximately 300. Hence if we measure Q in coulombs and E in volts, QE comes out in

$$\frac{\text{ergs}}{3 \times 10^9} \times 300, \text{ or in } \frac{\text{ergs}}{10^7}.$$

But 10^7 ergs makes one joule (p. 80); so that QE when measured in practical units comes out in joules. Also, the current I is equal to Q/t. Thus the work done per second by a current I is QE/t or IE, and is measured in joules per second, which we have already agreed to call watts (p. 81). Hence the power consumed in driving a current through a given potential difference is a number of watts equal to the product of the number of amperes by the number of volts. In a sort of scientific slang, we say ***amperes* $\times$ *volts* = *watts*.** It is one of the conveniences of the practical system of units that so simple a relation holds true.

From Ohm's law, since $I = E/R$, we may write

$$\text{Power} = IE = I^2R = \frac{E^2}{R},$$

and each of these expressions will be useful in practical calculations. The second one, for instance, tells us how much power is wasted in driving a given current through a given resistance, and shows that it is proportional not to the current itself, but to the square of the current.

Heat produced in a circuit. What becomes of the energy used to drive a current around a circuit? The answer depends on the nature of the circuit. If the circuit contains nothing but resistance, this energy is completely transformed into heat. Like the development of heat by friction this is one of the few 100% efficient processes in nature. The amount of heat thus obtained is known from the heat equivalent of work (p. 167). We have seen that 4.18 watts generate 1 calorie of heat per second. Thus a current of I amperes flowing through a resistance of R ohms generates $I^2R/4.18$ calories per second.

Electric heaters are much used nowadays. They are often marked according to the number of watts they require. Thus one marked 440 watts, to be used on a 110-volt circuit will generate 440/4.18, or 105.3 calories per second. It will require a current of $IE/E = 440/110 = 4$ amperes. An electric toaster consuming 5 amperes on the same circuit uses 550 watts and furnishes 131 calories per second.

It should be noted that the conclusions of this paragraph hold only if the circuit contains nothing but resistance. It might contain many other devices: storage cells being charged, in which energy was being put away in a chemical form; electric motors working elevators whereby weights were being raised; electromagnets, whereby some energy was being stored up in the form of a strong magnetic field; and so forth. In any such case, the power consumed in the circuit would, of course, be greater, usually much greater, than any heat that might be generated in the resistances that happen to be present.

Electric lighting. One of the most interesting uses to which the heat produced by a current in a resistance can be put is to generate light by this method, as in the common electric lamp. The light given out by a hot wire increases very rapidly with the temperature above 700° C. The wire passes from a dull red to a "white heat" at some much higher temperature, which might be estimated at 1500° C. or 2000° C. But the light becomes whiter (i.e., more like the daylight to which we are most accustomed) as the temperature goes still higher. In the history of the efforts to make light in this way we find that at first platinum wires were used in the manufacture of "incandescent lamps"; later carbon threads in a vacuum, and most recently fine wires of tungsten, either in a vacuum or better in an atmosphere of inactive gas. These improvements have all come about as steps in the direction of higher temperatures; the temperature of the tungsten wire in a nitrogen-filled lamp now reaches about 2800° C. The amount of light received per watt of power (now at the rate of less than half a watt per candle power in the larger lamps) has been increased at least six times by these changes, due to the fact that at higher temperatures a larger part of the energy goes into the short-wave part of the complete spectrum (p. 535). Still higher temperatures would be desirable, but the material of the wires evaporates from their surfaces at such a rate that the useful life of the lamp is unduly shortened. The "gas-filled" lamps are better than other incandescent lamps, because the inert gas in them gets in the way of the tungsten particles which would otherwise escape from the wires, and thus makes a long "life" possible at a higher temperature. The newest lamp uses mercury vapor for this purpose. It should be noted, however, that no incandescent lamp gives quite so much light for a given power expenditure as the arc does.

Electric arcs. High temperatures. If a 110-volt direct-current source is connected to a pair of carbon rods in contact, through a resistance which will hold the current down to 5 or 10 amperes, and then the rods are separated at the point of contact by a few millimeters, the gap is bridged over by an arc of violet glow, (Fig. 23–1), and the tip of each carbon terminal becomes very hot. Ordinary carbons made for this purpose are not very good conductors of heat, so that the tips reach a "white heat," the maximum temperature being limited, probably by the evaporation of the carbon rods. The hottest point of the positive carbon in air reaches and perhaps exceeds 3500° C. (6300° F.) while the other is some 500° C. cooler.

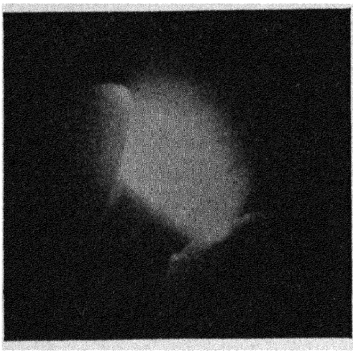


FIG. 23–1

A direct-current arc, showing the hot tips of the carbons, and the vapor between

The consumption of the carbon rods makes the maintenance of this form of arc somewhat expensive as a source of light, but it can be greatly reduced by enclosing the arc in an almost air-tight space, so that this fills up with the products of combustion rather than with air. The enclosed carbon arc is still somewhat used for street lighting and is provided with an ingenious automatic device which pushes the carbons forward as they burn away, and so keeps the length of the arc nearly constant. Other materials than carbon are also used in arc lamps. In the magnetite arc one terminal is a mixture of iron and titanium oxides compressed into a thin iron tube, while the other is of copper, which is hardly consumed at all. In this case the light comes from the flame rather than from the tips of the electrodes.

The arc may be used as a means of reaching very high temperatures. If an arc between large carbon rods (up to two feet in diameter) carrying a heavy current is buried in a mass of material which is non-conducting to heat as well as to electricity, we can produce in such an ***electric furnace*** an almost indefinite amount of heat and so reach a somewhat higher temperature (a little over 4000° C.) than in the arc directly. A limit to this temperature is set by the fact that the carbon evaporates, the material of the furnace melts and becomes conducting, and the heat diffuses out

from the place of concentration. An arc burning in carbon dioxide under a pressure of 30 atmospheres has been found to reach a temperature estimated to be well over 6000° C. In a vacuum electric furnace clean white sand is melted into transparent fused quartz, a most useful modern laboratory material, remarkable for the high temperatures it will stand and also for the suddenness with which its temperature may be changed without danger of cracking, due to its small coefficient of expansion. Such temperatures as exist at the surfaces of the stars (sometimes above 20,000° C.) cannot be produced in the laboratory except perhaps momentarily in explosions; and the temperatures which are estimated as existing in the interior of stars (over 1,000,000° C.) are as yet entirely beyond our reach.

Small arcs, formed in a stream of hydrogen gas, are now used for welding purposes. The flame is intensely hot, and at the same time the metal is kept clean by the hydrogen which prevents it from becoming oxidized.

Other applications of heat developed in resistances. Among the many practical applications which might be mentioned we select only a few. *Fuses* are wires of a soft, easily melted alloy made with different diameters, so adjusted that they will melt and automatically break an electric circuit when an excessive current is sent through them. They are much used in houses and laboratories where small currents are used. For breaking very large currents (100 amperes or so) magnetic circuit breakers are preferred. These consist of an electromagnet through which a part of the current flows, whose field becomes strong enough when the current increases to move a piece of iron and thus open a pair of contacts through which the current passes.

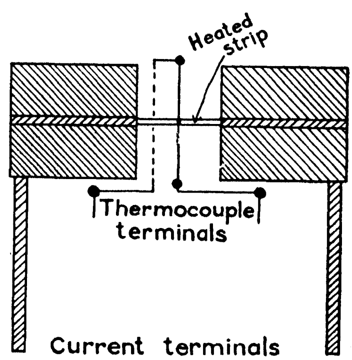


FIG. 23–2

Electric welding is a process in which the two rods, or other pieces of metal, to be welded together are placed in contact with each other, and then a large current is forced through them (see p. 424).

Hot-wire ammeters are instruments in which, in one form, the expansion of a heated wire is made to move a pointer over a calibrated scale, the divisions of which record the amount of the current warming the wire. Figure 23–2 shows a better form of ammeter, in which a strip, made of some metal which is not easily melted, is heated by a current, and a thermocouple fastened to the back of the strip indicates the temperature, and hence the amount of the current.

PROBLEMS

1. Find the time required to heat 500 cc. of water from 20° C. to 100° C., if the heat is developed in a 22-ohm resistance connected to a 110-volt supply. Assume that 80% of the heat generated is used, the rest lost to the surroundings.

2. A current of 0.5 ampere passes through an electric lamp with a fall of potential of 110 volts. Find the resistance of the lamp, the power consumed in running it, and the heat developed in it per second.

3. A building is furnished with current at a P. D. (at the building) of 110 volts, coming from a power house over a line whose resistance is 0.1 ohm. If 200 amperes of current are being used, find the fall of potential in the lines, the P. D. at the power house, and the power lost in the line.

4. An accurate method of measuring heat of evaporation is furnished by supplying heat electrically. If a 10-ohm heating coil is inserted in a vessel containing a liquid exhausted by an air pump, so that it evaporates at the rate of 6 grams per minute; and if 2 amperes of current must be sent through this resistance in order to keep the temperature constant while this is going on, what must be the heat of evaporation of this liquid, in calories per gram?

5. A trolley car line is supplied with 500 volts D. C. at the power house. If the car draws a current of 50 amperes over a line whose total resistance is 0.4 ohm, find the voltage at the trolley car, and the loss of power in the line from the power house to the car.

6. A 10-ampere current flows through a fully-charged storage battery of 0.05 ohm internal resistance. How much water evaporates per hour, if all the energy is used up in this way? (Assume 600 calories per gram.)

7. A current of 10 amperes passes through an electric arc across whose terminals there is a potential difference of 50 volts. Find the resistance of the arc, and the heat developed in it per second.

8. Why is an electric lamp spoken of as a 110-volt lamp? If one lamp has a resistance of 220 ohms, how much current would two such lamps draw from a 110-volt line when connected (*a*) in series, (*b*) in parallel? How many calories of heat would the lamp give out per second in cases (*a*) and (*b*)?

9. A large copper rod acting as one terminal of an arc is not wasted away, nor does its tip become white-hot as carbon does. Why not?

10. In the hot-wire ammeter, would you expect the scale to be uniform, or should its divisions be farther apart for high currents? Why?

11. Find the maximum power that can be produced by a dry cell of E. M. F. 1.5 volts, and internal resistance 0.1 ohm. What becomes of the energy in this case?

MAGNETISM AND CURRENTS

Electromagnets. It has already been noted (p. 333) that a long helical coil carrying a current acts like a magnet. The strength of

this magnet is enormously increased if a soft iron rod is inserted along the length of the coil, as in Fig. 23-3. It then becomes what we call an electromagnet. When the current is broken, the soft iron loses much of its magnetism at once, and more is lost if the rod is then jarred, heated, or struck by a hammer. Electromagnets have fields of force of the same form as permanent magnets but of much greater intensity. Physicists use the high magnetic fields thus produced in order to put matter under unusual conditions, and examine the changes thus brought about. It is found that

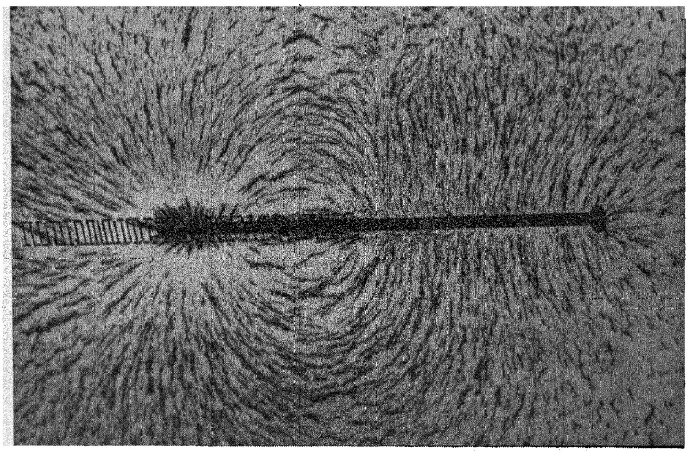


FIG. 23-3

A coil carrying a current (indicated by ink lines on the photograph) with a nail inserted part way as a core. The weak field at the open end of the coil is of no account compared with that produced by the part of the nail already inside.

magnetic materials change in length, others vary in electrical resistance, and several curious effects are produced on light by magnetic fields.

Electromagnets constitute the most noticeable parts of electric generators and motors; they are used to load and unload iron objects for shipment (as in Fig. 23-4), and to separate out magnetic ores. They are found in electric bells, door openers, telegraph instruments, electric organs, and a host of other appliances. The surgeon finds them handy for extracting small fragments of iron from an injured eye. The operation of most of these devices is easy to grasp if the fundamental physical ideas are clearly understood. With this aim in view we shall now return to a consideration of the magnetic properties of iron.

Magnetic properties of iron. The field about a magnet may be imagined to be filled with magnetic lines drawn in a quantitative way (p. 292) so that at any point in the field the intensity of the field is equal to the number of magnetic lines per square centimeter (the area being placed at right angles to the lines). These lines in the case of a uniformly magnetized bar are found coming out of the north pole and may be traced around till they disappear into the south pole. The phenomena of permeability, and of induction (p. 386) tempt us to regard them as continuous, but we need not attempt to follow them inside the iron, as no experiments can be conducted there.

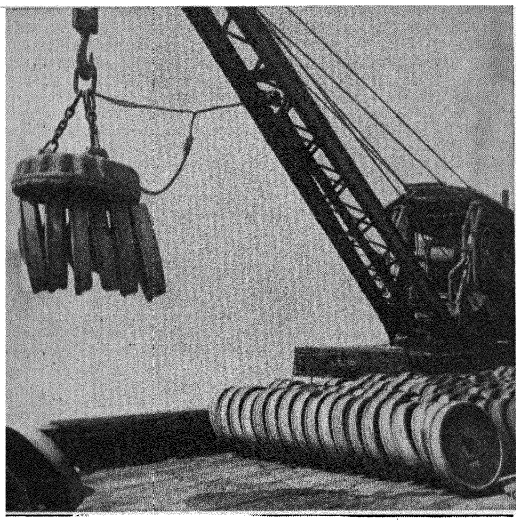


FIG. 23-4

This electromagnet has the shape (in cross-section) of a tall, narrow E which is hung up by the middle of its back, thus: m. The current flows in an inner, flat coil surrounding the central projection. The objects picked up thus serve to connect the central pole magnetically with the opposite pole, which takes the form of a ring surrounding it.

In a helical coil whose length is great compared with its diameter there is a certain field strength H which is known (p. 334). If the space in this coil is now filled up with a bar of soft iron, we find that it becomes a powerful electromagnet as soon as the current is turned on, and there are ways, using induced currents (p. 387), of finding how many lines of force emerge from the iron under these conditions. The number escaping from the ends per square centimeter is called the ***flux density*** and is indicated by the letter B. B is usually hundreds or thousands of times larger than H. The bar not only seems very "permeable" to the lines, but its presence greatly increases their number. If the iron bar is very long, so that its poles are so far off as to be thought of as exerting no considerable force inside the coil, then the magnetizing field may be assumed to be the same whether the iron is there or not, and H can be calculated, as above. The same is true if the iron

is in the form of a closed ring, with no poles.[1] The ratio B/H which can thus be found is called the ***permeability*** of the iron and is denoted by the letter μ ("mu"). Thus $\mu = B/H$. It is a quantitative expression of what we may roughly regard as the ease with which the lines are multiplied by the iron. We must, of course, remember that the lines may have no existence outside of our imaginations, and are useful only if they help us to describe the phenomena more compactly and remember them more easily.

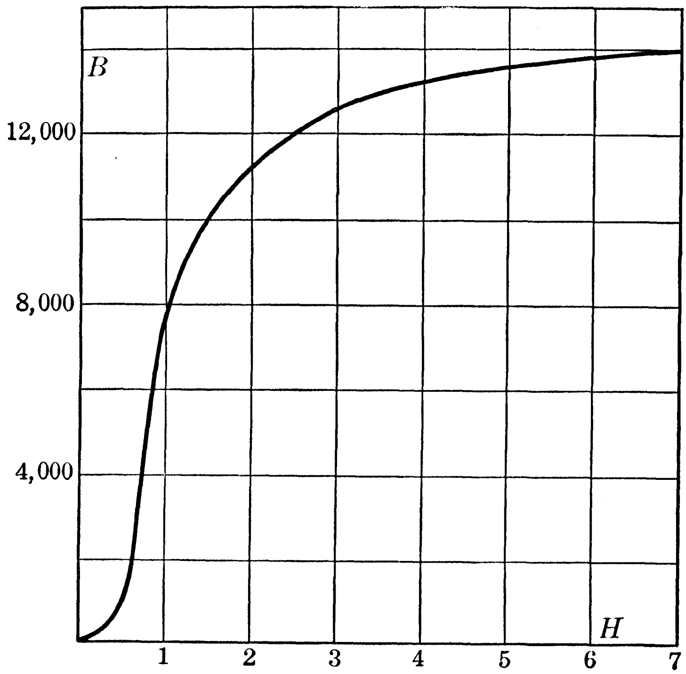


FIG. 23–5

A magnetization curve for soft iron

A typical B–H curve is shown in Fig. 23–5, where H is drawn on a very different scale from B. After passing the lowest values of H, there follows a steep, straight portion of the curve, until the "knee" is reached, after which it bends over and approaches saturation, further increase in the magnetizing current making very little change in the value of B. It is to be noted that soft iron yields a much higher value of B and is sooner saturated than hard steel.

[1] The measurement or calculation of H (or μ) in cases where the iron is in the form of a short bar, or where the magnetizing coil is not long, is a complicated matter which will not be considered here.

The shape of the B–H curve is just what one might expect if the atoms of the iron are little magnets themselves, oriented at random when the iron is unmagnetized. A certain strength of field may be needed to get them started, then they begin to face around in the same direction, and when they are all aligned in this way, the material is "saturated."

The way in which the permeability varies with the magnetizing field is shown in Fig. 23–6 in two cases. The largest value of μ

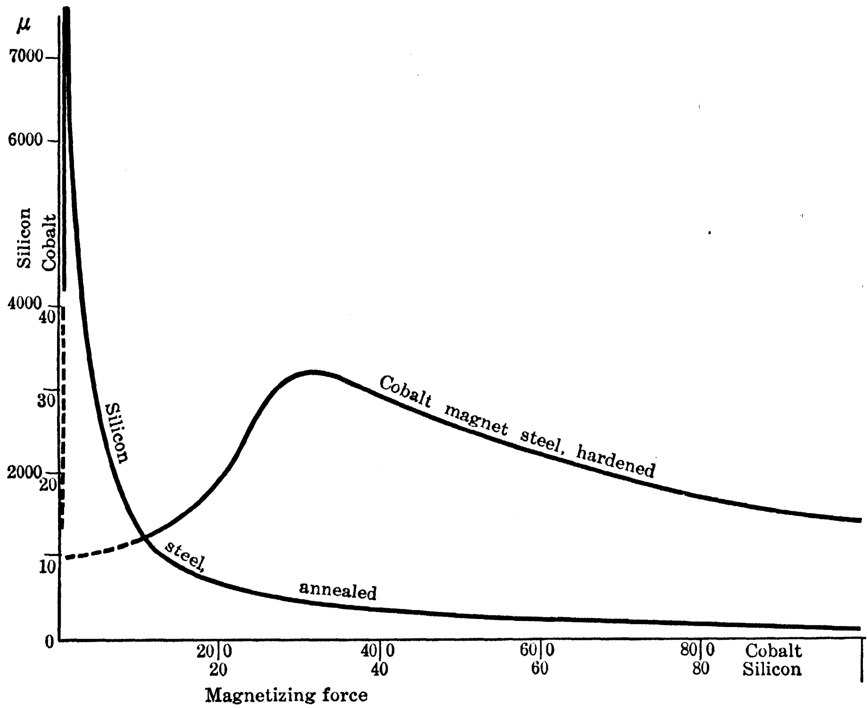


FIG. 23–6

corresponds to the steepest part of the B–H curve. It is easy to see that μ is far from being a constant quantity, at least for iron. It varies with the sort and condition of the iron, and with the current sent through the magnetizing coil. Still, even though it varies, it is a convenient term to use in practical cases where the field is to be calculated. It appears also in another connection which should be mentioned. If two magnetic poles of strengths m and m' and r cm. apart are immersed in a magnetic material of permeability μ, it can be shown that the force between them is less than usual; it should be $F = mm'/\mu r^2$. This is important in

electromagnetic theory, though experiments to test the truth of it are difficult, as the only available materials in which to try them are a few very feebly-magnetic liquids and gases.

If the magnetizing field is known to have H lines per square centimeter and we observe that it produces B lines per square centimeter coming out of the iron, we may say that $B = H + 4\pi I$, where I is called the *intensity of magnetization* of the iron. This is the same as the pole strength per square centimeter of pole face, and the 4π enters, since it can easily be shown that a unit pole has 4π lines in all coming out of it. The permeability can then be expressed by $\mu = B/H = 1 + 4\pi I/H$. The quantity I/H is called the *susceptibility*. Several other less used terms are found in the literature of this subject.

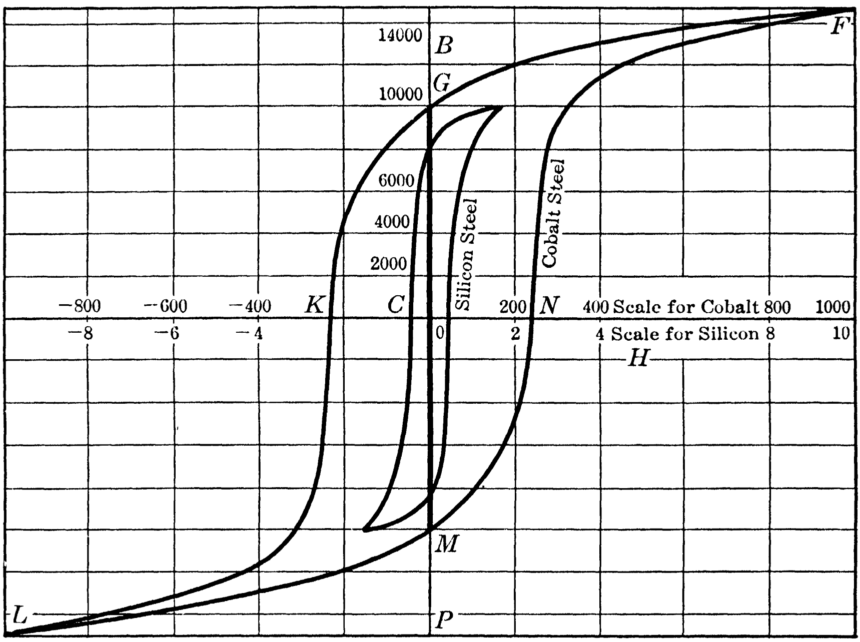


FIG. 23–7

Hysteresis curves for soft silicon steel and for cobalt steel on two scales. If the silicon steel were shown on the same scale as the cobalt steel, its curve would shrink so that it would all lie within the thickened central (vertical) line.

The magnetic cycle. Hysteresis. In alternating-current machines, which are very commonly used, the iron usually has to submit to constant repetitions of a ***magnetizing cycle.*** The magnetizing current, starting at zero in one direction, rises to a certain maximum, falls to zero, rises to an equal maximum in the opposite direction, returns to zero, and then repeats the same cycle over and over again. Figure 23–7 shows the response of the

iron. If we imagine it to start unmagnetized as at the point N, it rises to F, as the magnetizing force H increases. Then when the current falls to zero, the magnetism in the iron diminishes so that the flux density B reaches the value CG. It takes a small reverse field CK to reduce the first magnetization to zero. After this the reverse flux density rises to the value it has at L, drops to CM when the current again reaches zero, becomes zero again when the forward field reaches a value CN, and so completes the cycle. The magnetization of the iron must be thought of as tracing out such a loop for each alternation of the current. The descending and ascending paths on this B–H diagram do not coincide, as the figure shows, and this has an important physical result.

It can be shown that ***areas*** on the B–H diagram, which are proportional to $B \times H$, represent ***work done***. Hence the areas vertically under the curves CF and KGF in Fig. 23–7 represent respectively the work involved in two operations of magnetizing it to the state represented by the point F, and then demagnetizing it again. Corresponding areas in the lower left quadrant of the figure represent corresponding quantities with reference to the opposite magnetization. The sum total of all the work done in a cycle is represented by the area $LMNFGK$ included within the curve itself. Evidently this amount of work is done each time the iron is put through such a cycle, and this energy disappears into the interior of the iron in the form of heat. If this is inevitable and the operations are repeated many times a second, there may be generated in the iron a large amount of heat, perhaps enough to endanger the safety of the machine containing it. It is of great importance that this so-called ***hysteresis*** (meaning lag) loss should be accurately known for each sort of iron, and that some sorts should be found for which it is very small. The degree of success with which this has been done can be estimated later (see efficiency of transformers, p. 425).

The measurements of B and H which yield the hysteresis loop are too advanced to be described here. The area of the loop itself, after it is plotted, is easily obtained by means of a planimeter. Apparatus is sometimes arranged so that the loop is obtained quickly by a photographic method. A spot of light is made to move vertically in proportion to B and horizontally in proportion to H. The combined motion traces the loop if carried through a cycle.

Magnetic circuits. In electromagnets, and often in permanent magnets too, one sees that there is an almost complete circuit of iron provided, with only a small gap in which the function is performed for which the magnet is used. For instance, in an electrical generator (p. 406) there is a solid mass of iron through which the magnetic lines can pass most of the way around a complete circuit, and even in the gap, the space is largely occupied by the iron of the rotating armature. Similarly in the permanent magnet of the horse-shoe shape which is used in most ammeters and voltmeters the gap in the circuit is very small. This arrangement gives a highly permeable material for the lines to pass through almost all around their complete circuit, and in the electromagnet this leads to the production of many more lines than would otherwise be possible. The number of lines actually produced can in some cases be estimated by means of a helpful analogy which exists between magnetic and electric circuits.

In electric circuits the current is larger the more electromotive force there is, and smaller when the resistance is greater, according to Ohm's law. In magnetic circuits, the total number of lines, often called the "flux," ϕ, (though there is, of course, no such thing as a flow of magnetism, as this word would seem to imply), is larger the greater is the causative agent, here called the "magnetomotive force," M, and smaller when the hindering agent, here called the "reluctance," R, is greater. The relation among these quantities is then

$$\phi = \frac{M}{R}.$$

The magnetomotive force is caused by the current in the coil and must depend on the amount of the current (I) and the number (N) of turns in the magnetizing coils; or upon the number (NI) of "ampere-turns." The reluctance must increase with the length (L) of the magnetic circuit and decrease with increase of the area (A) of cross-section of the iron and the permeability (μ) of the iron. Thus R must be proportional to $L/\mu A$ and the flux is proportional to

$$\frac{NI}{\dfrac{L}{\mu A}}.$$

As a matter of fact the formula (proof omitted)

$$\phi = 0.4\pi NI \times \frac{\mu A}{L}$$

is approximately correct for many practical cases. If the circuit is not all of one sort of iron, or contains air gaps, the reluctance of the circuit is the sum of the reluctances of its different parts. For example, if the circuit contains one air gap (for which $\mu = 1$) and is otherwise made of iron of permeability 1000, the reluctance of this gap is 1000 times as large as though it were filled with iron. Thus the effect of an air gap in reducing the total flux is very large, and all such gaps should be made as short as possible if a large flux is to be produced.

Calculations of the flux to be expected with any electromagnet can be carried out, but they will not be considered here. They are subject to certain difficulties. For one, the permeability is not constant, but varies with the

same piece of iron if it is differently magnetized. If a suitable value for μ can be guessed at, there is still a considerable amount of "leakage," that is, some of the lines take longer paths, outside of the regions where they are supposed to lie. Nevertheless, such calculations are by no means worthless, as the leakage loss can be estimated by one familiar with the results of various designs of magnets, and thus commercially (if not scientifically) accurate results can be obtained.

Theories of magnetism. Ampère (1825) proposed the theory that in each atom showing magnetic properties these are caused by minute circular currents of electricity inside the atom. Our modern theory in one form replaces these currents by electrons revolving in orbits inside the atom (or something equivalent in its effects), which should produce the same magnetic fields as Ampère's currents, according to Rowland's experiment (p. 336). The natural reaction between these atomic magnets (often called "magnetons") and an outside magnetic field may produce two effects. One is to speed up or retard the rotation of the electrons in their orbits, and this can be shown to explain the phenomena of diamagnetism; the other is to orient the atoms, or molecules, of the substance so that they are all aligned in a definite manner. The latter effect will account for *paramagnetism* but it will occur only with those substances whose atoms have an exterior magnetic field, and by no means all atoms need be so built. It is quite possible (common, in fact) for the magnetons to occur in pairs or in groups within the atom so that all external effects are neutralized. Such atoms then can have no magnetic moment by means of which an external field can act on them, and they will show only a feeble *diamagnetic* effect.

These ideas were developed by Langevin (1905) and seem to be successful, with recent modifications, in explaining a wide range of magnetic phenomena.

Ampère's currents might be expected to show gyroscopic effects. In support of this idea, an experiment by Barnett should be mentioned in which a bar of iron was feebly magnetized just by spinning it, the spin tending to align the little atomic gyroscopes.

Magneto-striction. This is the name given to changes in the dimensions of bodies which occur when they are magnetized. The changes are all small, but such methods as are used in the laboratory for measuring heat expansion are delicate enough to show them. Iron and steel first lengthen, then shorten with increasing field strength; cast cobalt does exactly the opposite; nickel shortens at first rapidly, then more slowly. Recently this effect has been applied by Pierce [1] to the production of steady high-frequency oscillations in nickel rods which are likely to be of great use in radio telephony, and in other fields.

PROBLEMS

1. If an alternating current is sent through a coil with an iron core, how will the magnetic induction lag behind the magnetizing field on account of hysteresis? Will the induction follow a sine curve if the current does, or will its curve be distorted?

[1] G. W. Pierce, professor of physics, Harvard University, and director of the Cruft Laboratory.

2. A magnetic circuit consists of iron, 50 cm.^{2} in area, 80 cm. in length, of permeability 600, together with an adjustable air gap of the same cross-section as the iron. The magnet is wound with a coil of 400 turns, through which flows a current of 5 amperes. Find the gain in total flux if the air gap is reduced from 2 cm. length to 5 mm., assuming the flux formula given on p. 383.

3. A magnetic circuit consists of iron (length 50 cm., area 20 cm.^{2}, permeability 500) with an air gap (1 cm. long, 20 cm.^{2} area). It has a coil of 50 turns, carrying 2 amperes of current. Find the total number of lines (flux) in two cases, (*a*) as it is, and (*b*) when the gap is closed by a piece of the same sort of iron, just filling it.

4. Plot a curve showing how the permeability varies with the magnetizing field, given the following data:

B	H	B	H	B	H
40	0.32	1170	2.14	9,970	3.89
170	0.84	3710	2.67	11,640	4.50
420	1.37	7300	3.24	12,680	5.17

5. A new magnetic alloy, "permalloy," has been made, a bar of which will become saturated if placed parallel to the direction of the earth's total field intensity. How would a B–H curve for this alloy differ from that of an ordinary piece of iron?

CHAPTER 24

INDUCED CURRENTS

Simple experiments in induced currents, 386; induced E.M.F., 387; further experiments, 387; experiments with the earth's field; the earth inductor, 388; calculation of induced E.M.F., 389; direction of induced currents, law of Lenz, 390; eddy currents, 391; self-induction, 392; the induction coil, 394; the transformer, 396; the telephone, 396; the microphone, 397.

Simple experiments on induced currents. A new and remarkable source of electric currents was discovered by Faraday in 1831 and almost simultaneously by Henry.[1] Though Faraday's experiments astonished the scientific world of his day, nobody who witnessed them realized that they constituted the beginnings of an era in which electricity would become the servant of mankind. To reproduce some of his simplest results, we may use a small coil of a hundred turns, more or less, of insulated wire whose ends are connected to a fairly sensitive galvanometer, thus forming a circuit containing no source of E.M.F. If a fairly strong bar magnet is placed near it, a sudden motion of the galvanometer indicator will be observed each time the magnet is moved. If the magnet is thrust through the coil, a considerable current flows for an instant and then ceases, and on removing the magnet, an opposite rush of current is observed. Inserting a north pole produces a current in a direction opposite to that given by a south pole. A slow motion of the magnet produces a smaller effect than a quick one, though it lasts somewhat longer. Relative motion of the coil and the magnet is all that is required; the effects may all be obtained by keeping the magnet still and moving the coil.

[1] Joseph Henry (1797–1878), physicist, professor in the Albany Institute and Princeton University, later head of the Smithsonian Institution at Washington. He made an early form of telegraph, using silk covered wire for the first time in his coils; made a great lifting magnet to carry nearly two tons of iron; originated the idea of the relay; discovered the oscillations in a spark; showed that water has a force of cohesion equal to that of ice; organized a system of daily meteorological reports and weather maps, and made notable improvements in lighthouses. His originality led him to make worthy contributions to all the subjects he touched. His early work on induction makes it appropriate that the unit of inductance (p. 366) should be named after him.

When these experiments are examined, it is found that a current is obtained whenever the number of magnetic lines passing through ("threading") the coil is altered by the motion, and no current is produced if no such change happens. If a given number of lines are thus changed, doing so by a rapid motion makes a large current of very short duration, whereas a slow motion produces a small current which lasts throughout the longer duration of the motion. The current always ceases with the motion. In these cases the lines are "cut" by the coil, and the number threading it is thus changed. But it should be noted that in many cases (e.g. transformers) the coils cut no lines. The lines are created or perish within the coils; and yet the same effects are observed. On this account one should not limit the statement of induced currents to cases in which lines are cut.

These experiments are to be compared with the one done by Rowland (p. 336) in which it was shown that a moving charge produces a magnetic effect. Thus a charge and a magnet have no mutual action unless one of them is moved; moving the charge gives a current and thus a magnetic field; moving the magnet produces an electric effect, i.e., a temporary E.M.F. driving electrons in one direction or another. These effects are evidently two aspects of one action. They show that magnetism and electricity are closely linked, though as yet we have no simple explanation of how this is accomplished, or of the real nature of either of these phenomena.

Induced E.M.F. The current which flows in a metallic ring, or coil of wire, when a magnet is thrust into it, is due to an E.M.F. which is not localized in any one spot, but drives electrons forward in all parts of the coil. The threading of a single magnetic line appears to involve driving a definite number of electrons through a coil past any given point. Thus, a particular motion of a magnet causes the threading of a certain number of lines through the coil and drives a definite quantity of electricity through the coil circuit. The quantity of electricity involved can be shown to vary directly as the number of magnetic lines involved, and inversely as the resistance of the coil.

The E.M.F. creating such induced currents is comparable with the E.M.F. of a cell which would create the same current, and it is measured in the same units.

Further experiments. Two circuits can be arranged to show a

curious interaction when they are in no way connected, at least in no material way. One is made up of a coil and a galvanometer, as before, and the other of a similar coil in series with a dry cell and a key (Fig. 24–1). When the two coils are placed near together facing each other, there is a current induced in the one connected to the galvanometer when the key contact is made in the other circuit, and an opposite current induced when the key contact is broken. No current flows through the galvanometer when the current in the cell circuit is flowing steadily. The currents induced in this way are largest if the two coils are as near together as possible, and they are reversed either by inverting one of the coils, or by reversing the cell terminals. The occurrence of these currents can be described in terms similar to those used in the experiments with the magnet. Some of the magnetic lines created by the cell current in its coil thread the other coil and create the induced current.

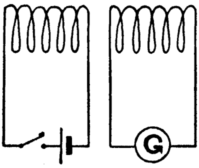


FIG. 24–1

There is a graphic way of looking at these phenomena which makes use of Faraday's lines in the magnetic field. If a current is just starting in a wire AB (Fig. 24–2) in the direction of the arrow, there will be circular magnetic lines growing about the wire and these will very soon reach a near-by wire CD, which may be thought of as capturing a little piece out of each line that passes. The line heals up again after the injury and goes on. As the figure shows, this gives rise to little temporary loops about CD in a sense opposite to those around AB. These collapse into the wire since they are supposed to be under tension, and their energy reappears in the form of the induced current. It is plain from the arrows that the induced current must be in a direction opposite to that in AB.

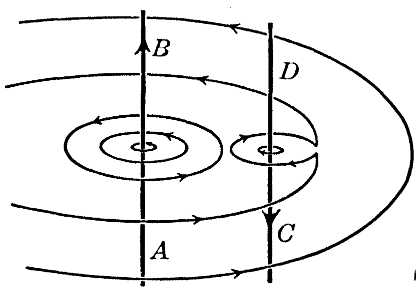


FIG. 24–2

This sort of picture is useful also in considering the field near a radio antenna which is emitting waves.

Experiments with the earth's field. Earth inductors. A coil, preferably of large size and of many turns, connected by long

flexible wires to a sensitive galvanometer, may produce considerable currents when it is made to cut the lines of the earth's field. Such coils are called "earth inductors." They are usually made with a spring mechanism which can be set to turn them over very quickly through a half turn. If the earth inductor is tried at various angles to the earth's field, the direction of the coil can be found at which it creates the greatest current when turned through 180°; i.e., when it cuts and re-admits the greatest number of lines in its motion. Its face must then be at right angles to these lines. This is seen to be true if one considers that, starting from this position, the first quarter turn cuts all the lines that passed through it at first, and the next quarter turn admits them all again, but in the opposite direction with respect to the coil. If we call the first direction of the lines positive, these are diminished to zero in the first quarter turn, and made negative in the second. Hence the change in the number of magnetic lines threading the coil is doubled in the half turn, compared with the change in either quarter.

When one finds the position of the coil for the greatest current, the line perpendicular to the coil must then be parallel to the earth's magnetic lines. From this the angle of dip can be obtained. This action is also the basis of one form of "earth-inductor compass," an instrument useful in aërial navigation.

Calculation of induced E.M.F. It is important to see how an induced E.M.F. can be calculated. In Fig. 24–3 *ABCD* is a frame of bare wire in a uniform magnetic field whose lines run perpendicularly to the diagram and are indicated, inside the frame only, by dots in the diagram. *FG* is a straight wire resting in contact with the frame and making a closed circuit *FBCG*. If *FG* is moved in the direction of the arrow, there will be a current induced in the circuit. This

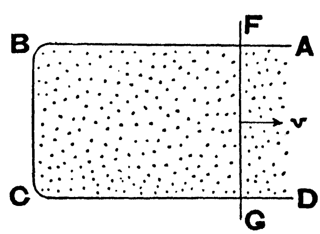


Fig. 24–3

current has energy in it. The principle of the conservation of energy shows us that there must be work done against electric or magnetic forces in moving *FG* and the amount of this work is just sufficient to supply the energy of the induced current. If work is done, the motion must be opposed by a force. This force, being caused by the field and by the current, must depend for its amount on both. We can see just how much force there is

by going through the following reasoning: the E–M unit of current flowing in a coil of 1 cm. radius produces a force on unit pole placed at the center which amounts to 1 dyne for each centimeter of the wire. By Newton's third law, this force is mutual; that is, the unit pole acts on each cm. of the wire with a force of 1 dyne. The wire being distant 1 cm. from a unit pole is by definition (p. 289) in a field of unit intensity. Thus, collecting all this together, we may say that each cm. of the wire carrying unit current in a unit field experiences a force of 1 dyne; or, L cm. of wire carrying I units of current in a field of B units intensity must experience a force of LIB dynes.

The work done per second in moving the wire across the magnetic lines is the product of the force by the distance traveled over per second, or Fv, if v is the velocity. But the work done is equal to IE (p. 372). Hence

$$IE = Fv = LIBv; \quad \text{or,} \quad E = LBv.$$

Since Lv is the area cut per second, LBv is the number of lines admitted per second into the loop, or, ***the E.M.F. in E–M units is equal to the rate at which the lines are being admitted per second*** (or the rate of change of the number threading the circuit); this value must be divided by 10^8 if we desire to have it in volts.

This result is true not only for the simple sort of circuit just treated but for all others in which the numbers of magnetic lines are in any way altered. If the circuit includes several loops or turns, as in an ordinary coil, the resulting E.M.F. is greater in proportion to the number of turns, or

$$E_{\text{volts}} = \frac{\text{rate of change of lines threading the coil} \times \text{number of turns}}{10^8}.$$

This formula applies also to motions which are not steady, the E.M.F. at any instant depending on the rate at which lines are then being cut. We shall find this conclusion useful in several cases.

Direction of induced currents. Law of Lenz. In the proof just given the principle of the conservation of energy required that work must be done in producing a motion against magnetic forces. The source of this work is the interaction between the magnetic field created by the induced current, and the magnetic field which is there already, and this action must always take the form of

opposition to the change which is occurring; otherwise the energy in the current will be obtained for nothing. This is the law first stated by Lenz[1] in 1833, and it is useful because it enables us to find the direction of the induced current in all cases.

For instance, the north pole of a magnet is approaching a coil. The current induced in the coil must produce a north magnetic face on the side of the coil which the pole is approaching, so that repulsion between the two (opposition to the motion) will occur. Hence the rule on page 333 tells us which way the current is flowing; it must go around the coil in a ***counter-clockwise*** direction, as seen from the magnet.

Similarly, if the north pole of the magnet is later withdrawn, the resulting induced current in the coil must make the coil equivalent to a south pole on the side of the magnet, pulling it back (opposing the motion); whence the current must be reversed as compared with the previous case, i.e., it must be clockwise.

In the case of the straight wire in Fig. 24–3, the current induced in the wire must produce a magnetic field which opposes the motion. The only way in which it can do this is by generating magnetic lines which point in the same direction as those of the field already there on the side toward which the motion occurs. Since Faraday imagined the lines to repel one another crosswise, this crowding of the lines ahead of the wire brings about the necessary opposition to the motion. Hence, when the directions of the main field and of the motion are specified, the direction of the current is known. If the lines of the main field are pointing inward (into the paper), the magnetic lines due to the current (in the case shown) go down into the paper on the right-hand side of FG (Fig. 24–3) and up on the left; which requires the current to flow from G to F.

Eddy currents. An interesting example of Lenz's law is furnished by the motion of a plate of copper (or aluminum) in a strongly concentrated magnetic field. If an attempt is made to move the plate rapidly through the field, strong "eddy" currents are produced, which move in curved paths around in the plate, and produce a field opposing the motion to such an extent as almost to stop it. If one holds the plate in one's hand, it feels as

[1] H. F. E. Lenz (1804–1865); physicist; professor in St. Petersburg (Leningrad). He made experiments especially in the field of electrical resistance, and the development of heat by currents.

though it were buried in thick molasses or some such viscous liquid. If the plate is slotted by a series of vertical cuts and then moved horizontally through the field (which is also horizontal), the effect is much diminished, as there is now no chance for the eddy currents to flow in any considerable amount.

Another striking illustration of such induced currents is shown in Fig. 24–4. A coil is wound about a core of soft iron wires, which projects a few inches above the end of the coil. A strong alternating current is sent through the coil, which produces rapid reversals in the magnetism of the core. A ring of aluminum surrounds the core, fitting loosely. Large currents are induced in this ring, which rush first in one direction, and then in the opposite, alternating with the frequency of the original current. In each direction the magnetic field which they create opposes the change that is occurring. Hence there is a force tending to drive the ring upward, and the photograph shows it supported near the top; actually, with this apparatus directly connected to a 110-volt source, the force is so great as to throw the ring several feet up into the air.

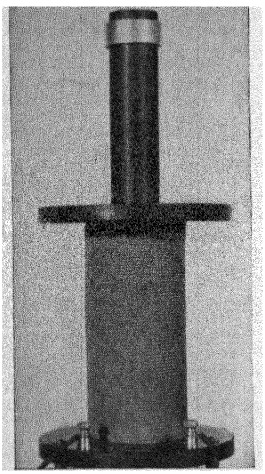


FIG. 24–4

Eddy currents are of frequent occurrence in the iron of magnetic machinery with moving parts, unless measures are taken to eliminate them. When they occur, they produce considerable losses of energy and cause the development of a great deal of heat in the iron. To avoid them, the iron is built up of sheets, or ***laminae,*** which are insulated by varnish or otherwise from one another. Laminated cores are, for instance, always used in transformers.

Eddy currents produced in a piece of metal may make it red-hot. In the ***induction furnace,*** a coil surrounds the specimen to be heated, and vigorous oscillations of current in the coil create heat in the metal in this way. This device is used to heat the metal parts of radio tubes after they are finished and exhausted, to melt conducting ores, or to melt non-conducting materials placed in graphite crucibles.

Self-induction. A case of induced electromotive force which is very important is the sort that occurs in a coil due to a changing

current in another part of the same coil or even in another part of the same wire. This effect is given the name of ***self-induction.*** Such coils are said to possess ***inductance.*** A growing current in any loop induces an ***opposing*** E. M. F. in adjacent loops, according to Lenz's law; a steady current does nothing; a decaying current induces a forward E. M. F. tending to prolong its own life.

This effect is large in coils which include a considerable area in each turn, or which have many turns, and it is enormously increased by the presence of an iron core in the coil. In circuits including the coils of large electromagnets, a high E.M.F. is induced which shows itself by slowing down the rate of growth of the current when the circuit is closed, and by producing a long arc which follows the blades of the opening switch when the current is being broken. It is impossible to make sudden changes of current in a circuit having large self-induction, and hence coils with iron cores (or sometimes without) are used to diminish fluctuations, or to prevent changes when steadiness of current is important. These are known as ***choke*** coils. One form of ***lightning arrester*** is so arranged that the sudden rushes of current caused by lightning are prevented by such a coil from going where they are not wanted, and, instead, they must jump a little gap and then proceed harmlessly along a line leading them into the earth.

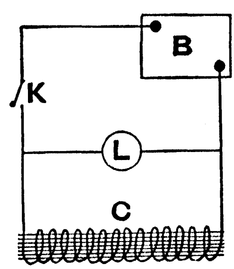


FIG. 24–5
A self-induction experiment

An easy way of exhibiting the effect of the high E. M. F. of self-induction when a circuit is broken is to arrange a circuit as in Fig. 24–5, in which B is a 6-volt storage battery, L is an 8-volt automobile lamp, and C is a coil with an iron core, as indicated by parallel lines drawn inside it, the conventional symbol for such a core. If the coil has a large self-induction, the first rush of current on closing the switch K will be through the lamp, but unless the lamp filament is very thin this will not show. After that the lamp will glow dimly, since the battery has not enough E. M. F. to make it bright; but, when the switch is opened, the E. M. F. induced in the coil at the break will drive a rush of current through the lamp, giving it a flash of brightness.

One may look at self-induction from the standpoint of energy also. When the circuit is first closed, the available energy of the battery is used for a short time in building up a strong magnetic

field in the iron core. No work is required to maintain a magnetic field, but the amount done in creating it is stored as energy in the field itself. Then when the current is broken, this energy cannot vanish instantly but must take another form. It produces the "extra-current" at the break, and causes the lamp to glow.

It is helpful to note that inductance in an electrical circuit acts in every way like inertia in mechanics. It prevents sudden starting or stopping of the current. When a current in a circuit is quickly broken, the effect is analogous to one in which a rapidly moving heavy body is stopped suddenly. In the latter case a large force is created by the inertia reaction (p. 67). In the electrical case, when a current in a coil with large self-induction is broken a high E. M. F. is produced.

The induction coil. A useful piece of apparatus for laboratory experiments is the ***induction coil.*** This device, though supplied with a current from a low-voltage (direct-current) source, such as a storage battery, furnishes intermittent jets of very high-voltage electricity, often at an E. M. F. many thousand times higher than that of the battery. In its action it is closely analogous to the hydraulic ram (p. 87). In both devices the inertia (electrical or mechanical) of a moving stream creates large effects (electromotive forces or pressures) when the current is suddenly checked. In the mechanical case this is easily accomplished. In the electrical case it is more difficult for the reason that when a switch is opened to stop the current, the phenomena of the electric arc or spark occur (ionization of the air, heat at the break, and vaporization of the material of the electrodes) and the current continues for some time by this sort of bridge across the widening gap in the switch. A special device has to be adopted to overcome this difficulty and make the break sudden.

The plan of the induction coil is indicated in Fig. 24–6.

The coil *PQ* consisting of a few turns of heavy wire is the primary coil through which current may be sent from the storage battery *B* through the contact point *F*, as shown. This coil has an iron core which is not solid but is made of a bundle of wires, so as to avoid eddy currents. When the core is magnetized, it pulls over the lump of iron *A* which is supported on a strip of spring brass, and this breaks the contact at *F*. Then the core loses most of its magnetism and the spring is again able to restore the contact, so that the process is repeated. The "interrupter" thus oscillates back and forth, like the hammer in an electric bell, continually starting and stopping the current in the primary coil. Each time at the "make" the growth of the current is quite

slow because of the self-induction of the primary coil with its core; at the "break," the decay would be delayed for the same reason and would be accompanied by sparking at the contact point, as explained above, with development of heat there and corrosion of the contact material, if it were not for the very large *condenser C* which is connected to opposite sides of this gap. This condenser consists usually of several hundred sheets of thin metallic foil, separated by paraffined paper, the whole being mounted compactly in the base of the instrument. The moment that the gap begins to open at *F* on the break, the high E. M. F. then induced by self-induction drives the current forward. There is a growing resistance in the widening gap, but an easy path into the condenser. The result is that the latter becomes highly charged, though in a very short time it discharges itself in the reverse direction through the primary coil, an operation which causes a quick demagnetization of the core, adding greatly to the E. M. F. induced in the secondary coil *S*. This coil may consist of miles of fine wire wound in thousands of turns around the primary. In each turn an induced E. M. F. arises when the number of lines of force threading it is changed, and as all the turns are in series, the resulting E. M. F. at the spark gap *G* may easily reach a value of hundreds of thousands of volts, so that a spark jumps across the gap, perhaps over a distance of many centimeters.

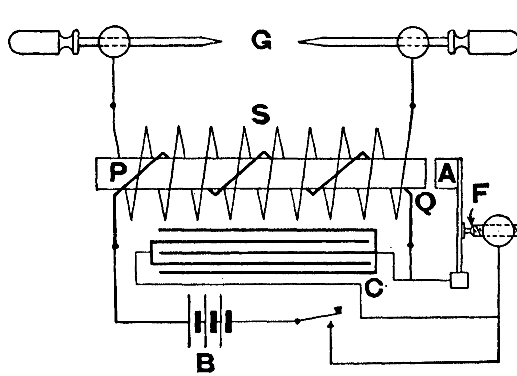


FIG. 24–6
The induction coil

There is a rough rule that about 8500 to 10,000 volts are required to make a spark jump between *sharp points* 1 cm. apart in air, and more in proportion for longer sparks, while if the sparks occur between *smooth surfaces*, such as balls, about 30,000 volts per centimeter are required. One may estimate the voltage obtained with an induction coil; thus a 4-inch (10 cm.) spark is given by an E. M. F. of the order of 100,000 volts, but the exact value varies greatly with the shape, size, and smoothness of the spark terminals, and the state of the air.

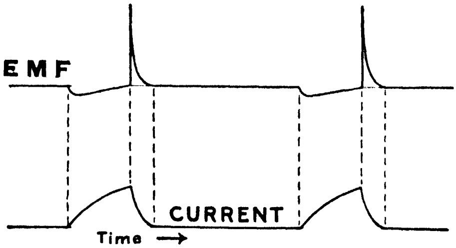


FIG. 24–7
The E. M. F. in the secondary coil, and the current existing at the same time in the primary

The high-potential discharge given by the induction coil occurs chiefly in one direction, as indicated graphically in Fig. 24–7, and this peculiarity enhances the usefulness of the instrument for certain purposes, such as the production of cathode

rays (p. 440). In the case of long sparks in air the aspect of the spark is a little different near the two terminals, the spark being brighter and seeming thicker near the negative terminal (see Fig. 20-7, p. 312); thus one can usually tell in which direction the current in the spark is flowing. If the gap between the spark terminals is short a spark may occur at the make as well as at the break. If this is not desired, a small additional spark gap may be put in series, and the discharge will then occur in one direction only.

The transformer. If primary and secondary coils are wound around a ***closed*** circuit of iron, instead of a straight core, or ***open*** magnetic circuit, we get another important device, the ***transformer.*** This, like the induction coil, may be used for the production of high voltages, but both the current supplied to this apparatus and that delivered by it are alternating rather than direct. It is discussed below (p. 423).

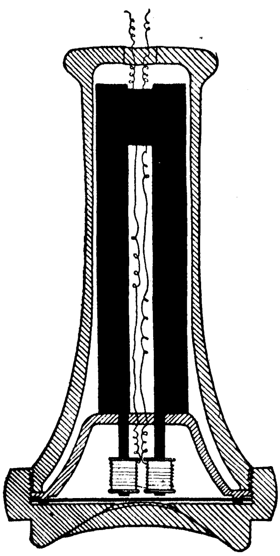


FIG. 24-8
A telephone receiver. The magnet is shown in black.

The telephone. In the telephone we have two parts, one a device whose function is to produce electrical oscillations corresponding to the air vibrations created by the voice, and the other a receiver which changes these oscillations back again into sound waves at the listener's ear.

The ***receiver*** contains a U-shaped magnet with coils about each pole (Fig. 24-8). The oscillations of current in these coils add to or take away from the magnetic field of the permanent magnet and thus alter the pull of the magnet upon a thin sheet, or diaphragm, of iron near by. This causes the iron to oscillate, and thus the vibrations are transmitted to the air outside. The modern telephone receiver is so made that an alternating current as small as a thousandth of an ampere is sufficient to produce ordinary conversational loudness, and one a thousand times smaller is still audible. The iron diaphragm has natural rates of vibration of its own, and responds more freely to these than to any others. This produces some "distortion" in the speech sounds, unduly strengthening certain frequencies at the expense of others, but this is not so great as to make the speech sounds unintelligible. On account of self-induction in the coils and inertia in the moving parts, the very

highest frequencies are not usually transmitted, and the lowest also are often absent. On the whole, however, the wonder is that the reproduction should be so good, rather than that there are defects. In recent years it has been possible to make great improvements in laboratory forms of telephones, though not all of these can be incorporated in commercial instruments.

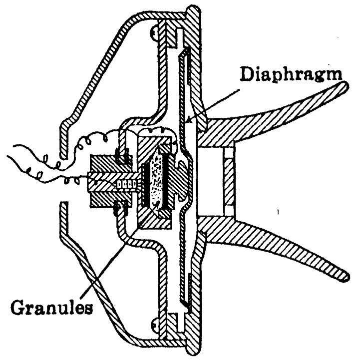


FIG. 24–9
A telephone ("carbon-button") transmitter

The microphone. The sending apparatus of the telephone involves a ***"carbon button" microphone,*** illustrated in section in Fig. 24–9. Figure 24–10 shows an experiment commonly used in illustration of the principle of the microphone. A light pointed carbon rod R rests vertically between carbon supports S and T, to which there are convenient electrical connections. The upper contact at R is a very light one, and in such cases, with carbon especially, the resistance varies considerably with the closeness of the contact, decreasing, of course, when the contact is good. A slight jar may be enough to make a large alteration in a current passing from a dry cell through S, R, and T, a change which will show on an ammeter included in the circuit. Thus a metronome ticking on the table near the carbon rod may make the ammeter needle jerk erratically from one reading to another.

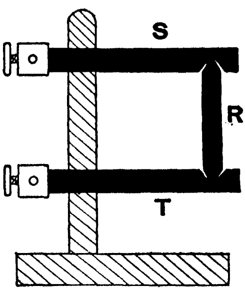


FIG. 24–10

If the carbon is in the form of granules between two conducting plates, as in the "carbon button," a more certain change is produced, compression always making a reduction of resistance, nearly in direct proportion. Thus the changes in air pressure due to voice vibrations in the telephone mouthpiece (Fig. 24–9) cause the diaphragm and hence one plate of the "button" to oscillate, producing variations in the resistance of the carbon between the two conducting plates of the button (shown in black), and hence in the current flowing in the circuit. These variations are transmitted over the telephone line to the receiver, either directly or

after they have been "stepped up" by means of an induction coil through whose primary coil they flow. The latter system is preferred for all but short distances. It raises the voltage and reduces the current in the lines, thus reducing the line losses and increasing the range over which messages can be sent. Telephonic "relays" have been devised which pass the vibrations along from one section of a long line to another without introducing any defects, and with renewed energy in each section. Thus, in telephoning over a distance of 1000 miles the original vibrations may travel over only a fifth of this distance, and the message may be relayed four times on the way without delay or confusion. The signals sent over a line are so much weakened ("attenuated") that even if current oscillations of thousands of amperes could be generated in New York they would not reach San Francisco on a direct line. The use of many relays offers the only possible means of success in long-distance telephone transmission.

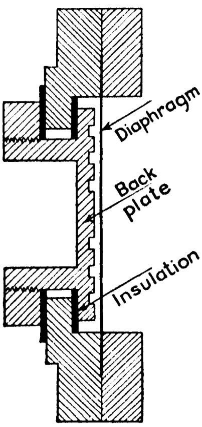


FIG. 24–11
Section of a condenser transmitter

The carbon microphone amplifies several times; i.e., there is more energy in the electrical oscillations which it produces than in the original sound vibrations. This additional energy is supplied by the battery from which the current flows. This microphone has been used successfully in the study of sounds under water, and in many other important experiments.

A ***double button,*** with carbon granules on each side of the diaphragm, is commonly used in radio broadcasting, giving a more faithful reproduction of sounds than occurs with the simpler type. The best apparatus, however, for distortionless reproduction is the ***condenser transmitter,*** in which one "plate" of a thin condenser is made of a light flexible metal diaphragm (Fig. 24–11). The motions of this plate which are caused by the sound waves alter the capacity of the condenser and thus cause electrical oscillations in the circuit.

PROBLEMS

1. In Fig. 24–3, if *FG* is 10 cm. long, and is moving at the rate of 3 m./sec. across a magnetic field of 5000 units intensity, what E. M. F. is being generated?

2. An electromagnet produces 6000 lines per square centimeter in a gap of 5 cm.^{2} area. A coil of 500 turns is placed around this gap in such a position

that all the lines thread the coil. It is then jerked out of the field, the operation taking 0.1 second. Find the average E. M. F. induced in the coil during this time.

3. A coil of 100 cm.^{2} and 200 turns is turned over by hand in a room where the earth's total magnetic intensity is 0.6 unit. If the turning takes place so that the maximum number of lines threads the coil in its first position, and again in its final position, what is the total number of lines cut per half revolution? If the coil is furnished with a commutator, so that direct current can be obtained from it, and if the coil revolves 10 times a second, what will be the average E. M. F. induced in the coil? Design a compass based on this idea. (Lindbergh used one in his flight across the Atlantic.)

4. An induction coil has a primary coil which produces 80,000 magnetic lines in its core when the current is flowing. If the secondary coil has 100,000 turns and the current is stopped in one two-hundredth of a second, what will the average E. M. F. in the secondary coil be during this time?

5. A coil of 10 turns of wire, area 10 cm.^{2}, is in a magnetic field of 10,000 lines per cm.^{2}, with its plane perpendicular to the field. If it is reversed so that the magnetic lines pass through it in the opposite sense, in 0.1 second, find the average E. M. F. in volts existing in the wire during the reversal.

CHAPTER 25

GENERATORS AND MOTORS

Action of simple generator. The transient nature of induced currents might at first sight make them appear unpromising for practical uses, where large electromotive forces must be continuously supplied. Yet, nearly all electrical power is now produced from such currents. We must examine the methods by which it is done.

A machine for creating current is called a ***generator.*** We shall begin by considering the action of a very simple type of generator, far too simple to be of much use, but easy to understand. It consists of a rectangular coil of several turns, AB, rotating about an axis C in a strong magnetic field, as shown in Fig. 25–1. Such a coil has a certain number of magnetic lines threading it at any instant, but this number is constantly changing as it rotates, and the E.M.F. induced in the coil is directly proportional to the rate of change of this number. An engine, or some other source of power, is required to make the coil rotate, and the coil must be connected with an external circuit if currents are to be derived from it.

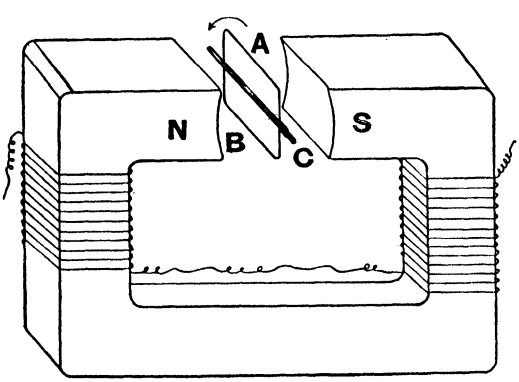


FIG. 25–1

If the coil is turning in the direction indicated in the figure, the wires on the side A will soon be cutting the lines next to the N

pole of the field magnet and therefore changing the number of lines threading the coil. According to Lenz's law there will be an opposition to this motion, and this must come from a repulsion of one set of magnetic lines on another. Therefore the current in A will produce magnetic lines running in the same (NS) direction ***below*** the wire as those of the field. To do this the magnetic lines due to the current in A must circulate ***counter-clockwise*** according to the screwdriver rule on page 333, or the current in A must be approaching the reader.[1] In like manner the lines will go ***clockwise*** around B, and the current in that part of the coil will recede from the reader. The other two sides of the rectangular coil are moving

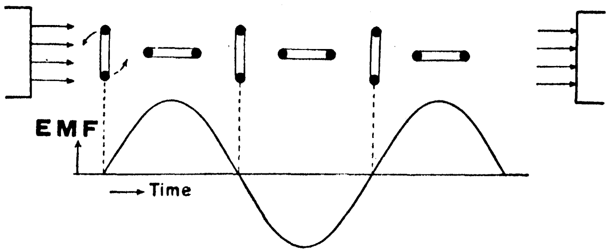


FIG. 25–2

Above is shown the coil in several positions in the magnetic field. Opposite each (below) is the E. M. F. being induced.

parallel to the magnetic lines of the field, and therefore cut no lines in their motion. Thus, so far, we find that this motion generates an E.M.F. causing a current to circulate about the coil in one sense. Next, when A has reached the lowest point in its revolution, it is for the moment cutting no lines of force, and the E.M.F. in it has become zero. As A begins to rise on the right-hand side, the current goes in it as it formerly went in B, when B was in that position. With reference to the wire in A itself the current now goes in a direction opposite to that in which it went at first. Thus inside the coil itself the E.M.F. and the current are alternating ones, reversing each time that the coil becomes vertical (shown in a schematic way in Fig. 25–2).

[1] In treating the directions of currents and magnetic lines many practical people prefer to use Fleming's *right- and left-hand rules*. The right-hand rule applies to generators and is as follows: with thumb, first finger and middle finger extended so as to be at right angles to one another, point the thumb in the direction of motion of the wire, the first finger in the direction of the magnetic lines, and then the middle finger will show the direction of the induced E.M.F. or current.

So far we have not shown any outer wires connected to the coil AB. If we now connect each end of this coil to an insulated metal "slip ring" mounted on the same shaft and rotating with the coil (Fig. 25–3), and then place two stationary metallic (or carbon) "brushes," so that they will make a rubbing contact with the rings as they rotate, we shall be able to draw off through these brushes an alternating current, which will flow through an external circuit. Alternating currents are considered more fully in the next chapter.

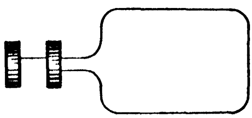


Fig. 25–3

If instead of a pair of solid slip rings the ends of the rotating coil are connected to the two halves of a split ring (Fig. 25–4a) against opposite sides of which both brushes make contact at once (Fig. 25–4b), then we can draw a current from the coil which flows through the external circuit always in the same direction. As this split ring (or "commutator") rotates, the brush B_1 makes contact first with one half (i.e., with one end of the coil) and then, half a turn later, with the opposite half. If the split in the commutator passes the brushes at the right time (which can be adjusted by turning the holder supporting the brushes shown in Fig. 25–11), this transfer from one to the other half of the split ring occurs at the moment when no current is being generated. Then the current drawn from the brushes through the external circuit is of the form shown in Fig. 25–5, that is, it is a pulsating direct current, with two surges to each revolution of the coil. Such a varying current would not be generally useful, though it might do well enough in a few cases; for instance, for charging storage batteries.

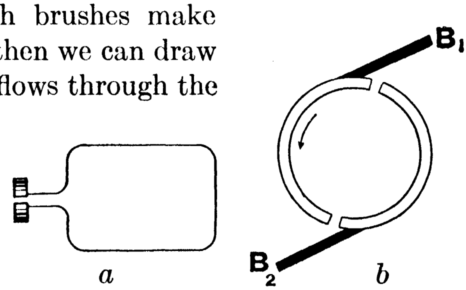


Fig. 25–4

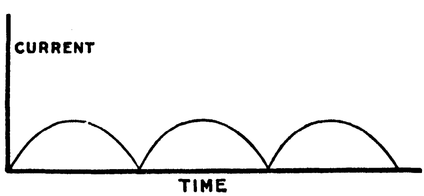


Fig. 25–5

A steadier current could be produced if there were two coils, each generating current, set at right angles to each other. When

one was inactive, the other would be in its best position. The current, as in Fig. 25–6, would then never drop to zero, and would rise and fall through a narrow range. Evidently it would be still better to have many coils, set at regular angular intervals.

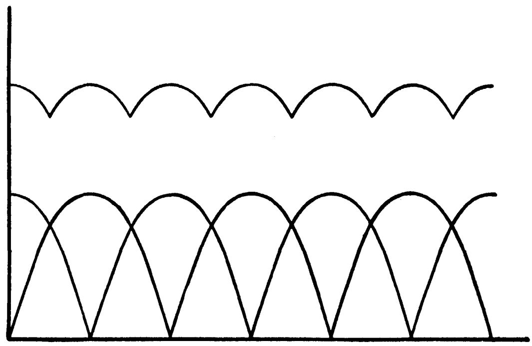


Fig. 25–6

Generator with many coils. The rotating part of a generator such as we are considering is called its ***armature.*** Its body, or core, is made of iron in order to lower the reluctance and increase the number of magnetic lines in it. The coils are wound in slots cut in the iron, thereby reducing the air gap. Figure 25–7 shows an armature wound with many coils, each containing several turns. The particular arrangement shown in the figure is called "drum winding"; each coil is a rectangular one of the shape shown in Fig. 25–3, but the coils form a continuous circuit; that is, the end of one coil is connected to the beginning of the next all the way around. The ends of each coil are also connected to narrow metallic segments in a very much split commutator, with as many segments as there are coils. The brushes bear against this commutator as they did in the case of the simple generator.

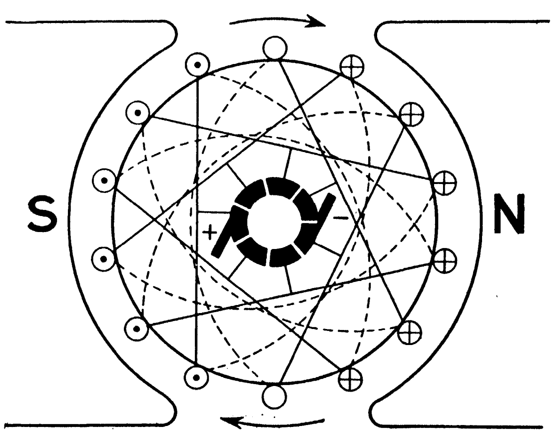


Fig. 25–7

To follow out the action of this device, let us suppose that the wires on one side of the armature are generating E.M.F. in such a direction that the current flows to the upper brush and so out to the external circuit; that is, this brush is the positive one. It will

be seen, on examining the direction of the motion and of the magnetic lines, that on the opposite side of the armature the current is also flowing to the upper brush. The polarity of the two sides agrees, and the wires on each side resemble a number of cells connected in series, delivering current to the brushes through the segments which they happen to be touching at the moment. As the armature turns, this condition is maintained unchanged.

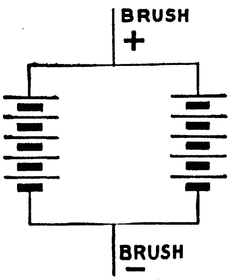


Fig. 25–8

Figure 25–8 shows schematically how two groups of cells could be arranged to imitate the action of the windings of the coils. The cells are in two groups in parallel, the positive terminal being similarly situated in each.

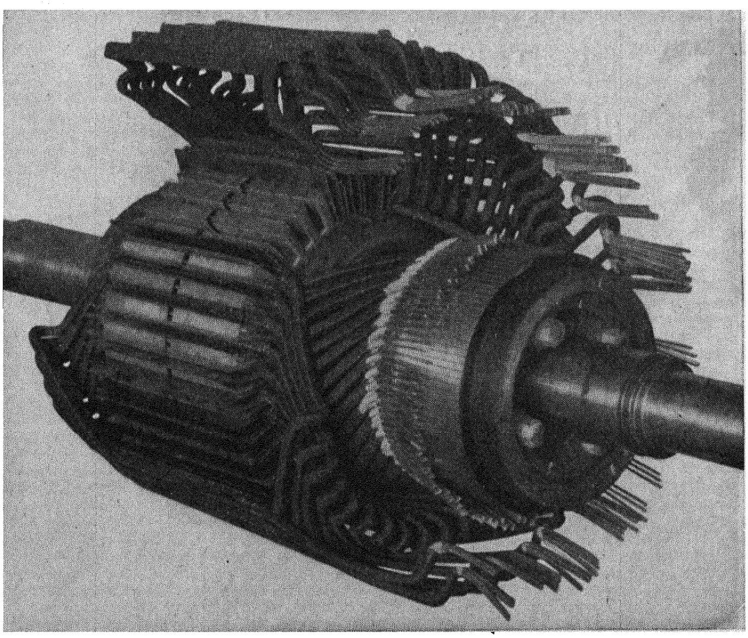


Fig. 25–9
A direct-current armature partly assembled. (Courtesy of the General Electric Company)

Figure 25–9 shows an actual armature in an unfinished state. There are empty slots and others with one or more coils in place. Other coils, previously "formed," are shown ready to

be put into place. At the near end is the split commutator, which in this case is divided into about a hundred parts.

If the number of segments in the commutator is large, the E.M.F. obtained from such a direct-current generator is very nearly steady, there being left only traces of the former pulsations, which now form a "commutator ripple" (Fig. 25–10). This is usually unimportant; its effect can often be heard as a musical note in a direct-current arc fed by the machine.

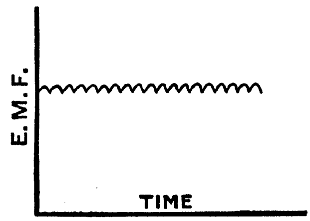


FIG. 25–10
A commutator ripple in the E. M. F. (greatly exaggerated)

Figure 25–11 shows a four-pole, direct-current machine (fitted also with four extra "commutating" poles, whose function is to

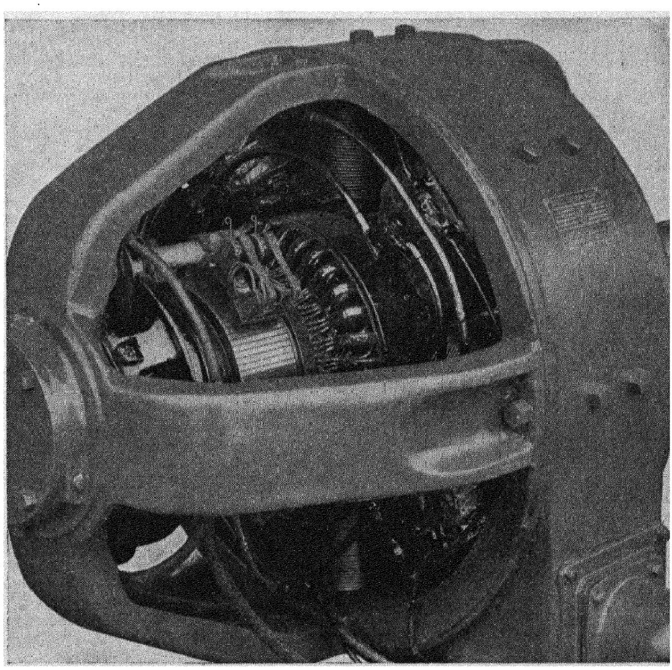


FIG. 25–11
A direct-current generator. (Courtesy of the General Electric Company)

improve commutation). One of the four sets of brushes is plainly seen resting against the split commutator.

The magnetic circuit of a generator. It is to be noted that the electromagnets (called the field magnets) which produce the field

form with the iron in the armature a complete iron circuit, broken only by two air gaps, in which the coils have barely room enough to rotate. (Figure 25–12 shows the magnetic circuit of a four-pole machine.) In this way the reluctance (p. 383) is reduced to a minimum, and the total number of lines (the flux) is made as large as possible. The iron core of the armature is laminated (p. 392) to prevent eddy currents.

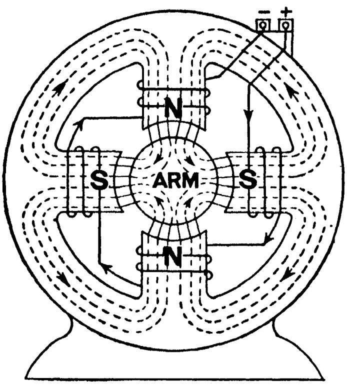


Fig. 25–12
The magnetic circuit of a four-pole generator

The current for the field coils is derived from the armature itself, as the next paragraph shows. The residual magnetism in the magnetic circuit is always enough to begin with, and the field is gradually built up as the speed increases.

Series, shunt, and compound windings. The field-magnet windings can be connected in any one of three ways. If they are in ***series*** with the external circuit and the armature (Fig. 25–13)

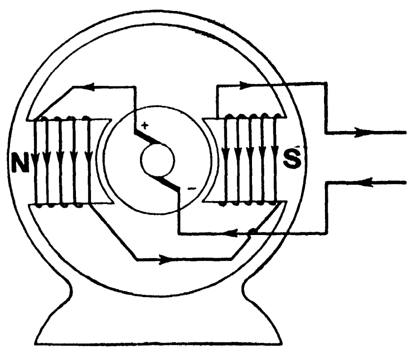


Fig. 25–13
A series-wound generator

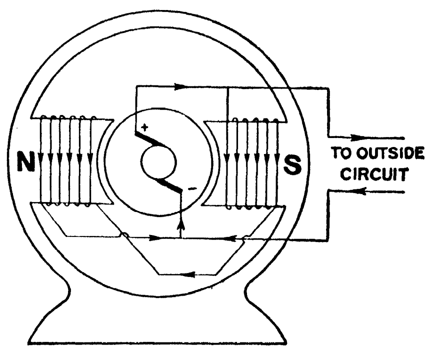


Fig. 25–14
A shunt-wound generator

all the current that flows in the external circuit flows through the field coils also. It follows that when the current from the generator increases, the field magnets become stronger, there are more lines to be cut, and, since the machine is kept running at constant speed, the E.M.F. of the machine rises. But, when the iron of the electromagnet becomes saturated this rise stops, and the terminal voltage then falls rapidly with a further increase in

current because of the *IR* drop in the machine itself. The change of terminal voltage with current is shown in Fig. 25–16. These series generators are used in direct-current street arc lighting, and are then operated on the descending portion of their "characteristic" curve, throughout which they furnish an almost constant current.

If the field coils are in ***shunt*** (or ***parallel***) with the outer circuit, as shown in Fig. 25–14, a large current drawn from the machine lowers the difference of potential between the points (the ends of the armature windings) to which the field coils are connected, and hence the current in these coils drops and the terminal voltage of the machine decreases somewhat, as shown in Fig. 25–16. This de-

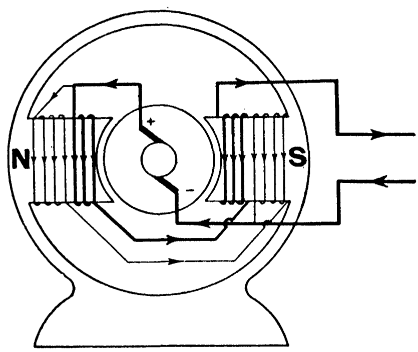


Fig. 25–15
A compound-wound generator

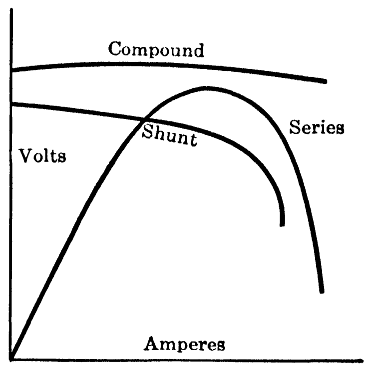


Fig. 25–16
Characteristic curves of three types of generators

crease can be prevented by regulating the current in the field coils by auxiliary means. Such machines are used in lighting systems.

A more uniform terminal voltage can be derived from a machine which is ***compound-wound;*** that is, it has some windings of its field coils in series with the external circuit and the rest of them in parallel thereto (Fig. 25–15). Within limits, the terminal voltage given by such a generator may be made independent of the external current, which is a desirable characteristic if the machine is used for laboratory experiments or to light a varying number of incandescent lamps. Figure 25–16 shows the characteristic curve of this type also; i.e., the variations in terminal voltage at constant speed when the load is increased.

Multipolar generators. It is common practice to build generators with more than two poles. Figure 25–11 shows the form of a

four-pole generator and Fig. 25–12 shows the arrangement of its magnetic fields. The armature windings, revolving close to the poles, cut in one revolution all the magnetic lines, of which there may now be nearly twice as many as there were with a two-pole machine. Hence, to get the same voltage a lower speed is sufficient. There is also some economy in the weight of iron needed.

Calculation of the E.M.F. of a generator. The general rule given on p. 390 states that the E.M.F. (in volts) induced in a coil of n turns when the number of magnetic lines threading the coil is changing is equal to ***n times the rate of change of the number of lines***, divided by 10^8. In a two-pole generator, each half of the armature generates the same voltage, and these two are in parallel, as shown in Fig. 25–8, p. 404. Hence, the proper number to use for n is really the number of windings, or individual turns, the opposite sides of the same turn not counting double.

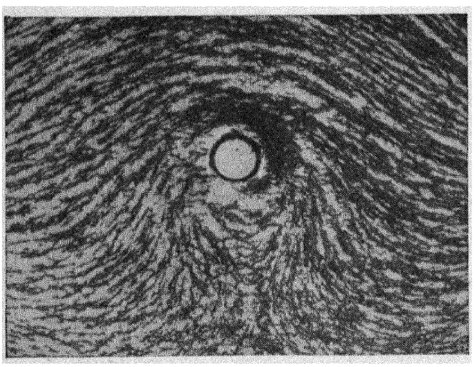


FIG. 25–17

The voltage must evidently vary directly with the speed if the field remains constant. If, on the other hand, the speed is kept constant and the current in the field coils is varied, the E.M.F. generated will change, following a curve like the B–H magnetization curve, Fig. 23–5, p. 379, as the strength of the field alters.

Direct-current motors. The same machine that is used as a generator may also be used as a motor. In this case the current enters from an outside source, through the brushes and the split commutator, so as to pass through the left-hand wires in the armature always in the same direction. When these by their rotation become right-hand wires, the current in them is at the same moment reversed by the action of the commutator. The current through a coil in a magnetic field must produce a torque on the coil. There is a force F acting on a single conductor of length L when a current of I E–M units flows through it in a magnetic field of intensity B, and the magnitude of this force is given by the formula $F = BIL$ (p. 390). Knowing the dimensions of the coil, the

torque produced by this force on each conductor in the coil, and hence on the coil as a whole, can be found.

The direction of the resulting motion of the coil can readily be determined. The screw-driver rule (p. 333) gives the direction of the magnetic lines due to the current. On one side of each conductor these run in the same direction as the lines due to the field itself, and the lines are crowded there. This crowding produces repulsion and drives the wire away from that region.[1] Figure 25–17 shows (by means of iron filings) how this happens.

Back E.M.F. in motors. When a motor is running, its coils are cutting magnetic lines, and it is therefore at the same time generating an E.M.F. Lenz's law tells us that the E.M.F. which it generates must be in opposition to the entering current; hence it is called "back E.M.F." It must increase with the speed, and will rise to a high value when the machine is running rapidly. Hence the current at first will be very much larger than it is later. If the full-speed current is as large as the motor can stand without overheating, the current at starting will create far too much heat in the armature wires, and may injure their insulation. With very small motors, the armature is so light that it attains full speed before any damage can be done; with large motors, some resistance must be put in the circuit at first, to hold the current down to a safe value, and this resistance is reduced as the motor gains speed. This operation is familiar to anyone who has watched a motorman turn his handle from notch to notch in starting a street car. In the case of a machine-shop motor, the "starting-box" mounted near it has similar steps in its resistance and often has an electromagnet to hold its switch "on" at full speed, and a spring to pull it "off" when the current ceases.

Action of different types of D.C. motors. Direct-current motors may be wound with their field coils in series with the armature, in shunt, or "compounded," as in the case of generators.

The series-wound motor is used in street cars, electric locomotives, and automobile starters, where a powerful torque is required at starting. In this form the torque depends directly on the cur-

[1] Fleming's *left-hand rule* also gives the direction of rotation for *motors.* It is similar in wording to the right-hand rule (footnote, p. 401). The left thumb indicates the direction of motion of the wire, if the first finger points along the field lines and the middle finger (at right angles to the other two) in the direction of the current.

rent in the armature; and the field varies (roughly) with the current in the field coils. Since the same current goes through both, the torque varies nearly as the square of the current, which makes the motor unusually powerful at first, when a large current is sent through it, and its speed is still low. Such motors are geared to their load, rather than belted, as they would race dangerously if the belt were suddenly broken.

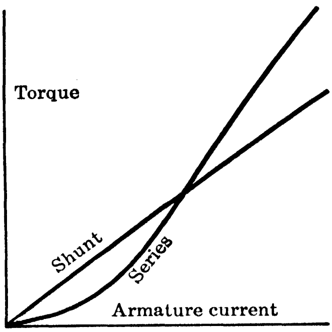


FIG. 25–18
Characteristic curves of motors

The shunt-wound motor is steadier in its speed and gives a torque proportional to the current. It is used when a constant speed is required, as for instance in blowers, or in a machine shop.

The compound-wound motor is rarely used except for elevators, punch presses, etc. where a large initial torque is demanded; and then, in order to keep it from speeding up after starting, the series part of the field winding is automatically cut out when the machine reaches a certain speed.

The curves in Fig. 25–18 illustrate the performance of the shunt and series types of motors, giving the torque produced for each value of the armature current.

PROBLEMS

1. Is it proper to speak of a generator as a source of electricity? Where is the electricity before it is "generated"? What does a generator generate?

2. Electric locomotives use their motors instead of brakes to hold them back on long down grades. Explain how this can be done, and how the energy might be disposed of.

3. A vacuum-cleaner motor is attached directly to a 110-volt D. C. circuit, and takes 1 ampere of current through its armature, and a negligible amount through its field coils. If the resistance of the motor is 2 ohms (*a*) what must be the back E. M. F. of the motor when running at rated speed, (*b*) what current would flow if the armature could not rotate, (*c*) how many calories would this current produce per second in the motor?

4. A simple 2-pole D. C. generator runs at a speed of 600 revolutions per minute, and its armature contains 100 turns. If the total number of lines cut by each turn in passing in front of each pole is 600,000, find the voltage generated by the machine. What effect might a change (*a*) in speed, (*b*) in the number of turns on the armature, and (*c*) in the magnetizing current in the coils of the field magnets have on this voltage?

5. A motor whose armature has a resistance of 1 ohm runs on a 110-volt circuit. It draws 5 amperes when running at rated speed, and this is all the current that ought to be allowed to flow through it. (*a*) Find the back E. M. F. at rated speed. (*b*) Find the greatest resistance which should be put in series with the motor in its "starting box," when it is being started from rest.

6. At what speed must a 110-volt D. C. generator with drum armature be run if it has 2 poles with 1,000,000 magnetic lines crossing from one to the other, and if the armature has 440 turns?

7. What is the average E. M. F. induced in an armature which has 100 conductors in series, and rotates so that these conductors cut across a field of 5,000,000 lines 800 times a minute?

8. A shunt motor is connected to a 115-volt line. The field windings have a resistance of 200 ohms, and the armature a resistance of 2 ohms. (*a*) What current will it take from the line when the machine is at rest? (*b*) What is its back E. M. F. when running at rated speed if it then takes 3.2 amperes?

9. Find the efficiency of a certain motor from the following data, given on its name plate: 5 H. P.; 110 volts; 38.5 amperes; 1800 R. P. M. (746 watts = 1 H. P.).

10. A D. C. motor runs fast enough to generate a back E. M. F. of 100 volts when connected to a 110-volt source. It then takes 5 amperes of current through its armature. What current would it take if it were connected directly to the 110-volt source, the armature not being permitted to revolve?

11. A D. C. motor has an armature resistance of 1 ohm. When connected to 110-volt source and running at a certain speed, it takes a current of 10 amperes. Find the back E. M. F. at this speed, and also the current which it would take if running at half this speed.

12. (*a*) Find the back E. M. F. in a shunt D. C. motor if the armature runs at a constant speed; armature resistance 0.24 ohms; impressed voltage 112 volts, and armature current 24 amperes.

(*b*) If the speed were reduced one-half, and the field remained the same, what current would flow in the armature?

(*c*) If the armature were stopped altogether, what current would flow in it?

13. The output of a generator running at 1200 revolutions per minute is 50 amperes at 100 volts. Find the number of kilowatts required to run it, (assuming no friction) and the torque (in grams-cm.) needed to turn it.

14. The resistance of a series-wound motor running on 110-volts D. C. is 2 ohms, and its back E. M. F. at its normal speed is 104 volts. Find the maximum current at normal speed, the heat developed by this current in the machine, and the power produced (assuming no friction). Find also the resistance that must be connected in series in starting, so that the heat developed by the current in the machine will not be more than 4 times as great as that produced at normal speed.

CHAPTER 26

ALTERNATING CURRENTS

The nature and measurement of alternating currents, 412; phase lag due to inductance, 413; phase change due to capacity, 414; impedance, 415; impedance of circuits containing resistance, inductance and capacity, 416; power in A.C. circuits; power factor, 418; alternating current resistance; skin effect, 419; alternating current generators, 420; transformers, 423; step-down transformer, 424; electrical welding, 424; power loss in transformers, 425; power transmission, 425; rotary converters, 426; rotary fields; two-phase currents, 426; two-phase induction motor, 427; single-phase motors, 428; three-phase currents, 428; motors for A.C. or D.C., 429; A.C. ammeters, voltmeters, and wattmeters, 429.

The nature and measurement of alternating currents. We have already seen (p. 402) how easily an alternating current is generated and how the current flows back and forth through a circuit (Fig. 25–2, p. 401). We find that in practice alternating currents are used almost exclusively for electric lighting and for the transmission of power. It will be worth while to examine more fully the nature of these currents.

A difficulty presents itself when we come to measure them. Evidently the average current over a complete cycle is zero, no matter how large a value the current may have at any instant. Should we then measure the current by its maximum value or according to some other scheme? The sensible answer is to consider the energy in the current, and from this to derive a definite meaning for "one ampere" of alternating current. If a direct current I flows through a resistance, the heat developed is proportional to I^2 and is independent of the direction of flow. This suggests that the sensible way to measure an alternating current is by finding the square root of the average value of I^2 through a complete cycle. This is known as the "R. M. S." (root mean square) value of I, or sometimes as the "effective" value of the current, or its value in "virtual amperes."

In like manner we measure an alternating E. M. F. in R. M. S. (or "virtual") volts.

The average value of the square of a quantity which fluctuates according to a simple periodic curve like that in Fig. 25–2 is easily found by calculus methods. ***The maximum value*** is found to be equal to $\sqrt{2}$ times the R.M.S. value; or, the R.M.S. value is 0.707 times the maximum. Thus 1 ampere of alternating current ("1 A.C. ampere") varies from a maximum of 1.414 amperes to zero, positive and negative.

Phase lag due to inductance. An alternating current in a circuit which contains inductance has some peculiarities which are easily approached by first considering an analogous mechanical arrangement. On a frictionless level surface we imagine a box *B* to be placed, as in Fig. 26–1, in which there is a heavy mass *M*. The mass *M* is to be moved back and forth by hand with simple periodic motion through a small distance. The force needed to do this must be different in amount at each point in the oscillation. It is greatest at the ends of the motion where the acceleration of the mass is greatest, and where the inertia reaction (p. 67) is large. The greatest push to the right must be

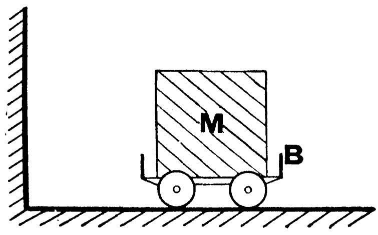


FIG. 26–1
A mechanical analogue of an inductance

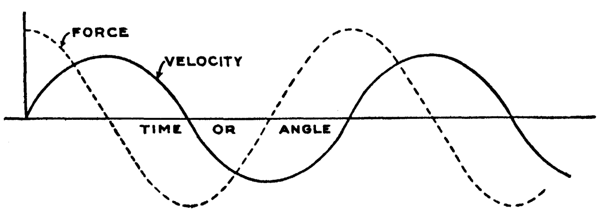


FIG. 26–2
Force and velocity in the mechanical model, or E. M. F. and current in a circuit containing inductance only

exerted on the mass when it is farthest to the left, but the greatest velocity to the right does not occur until a quarter of a period later, when the mass is in the middle of its swing. If we should draw the force f and the velocity v on the same diagram, as in Fig. 26–2 (on different scales, of course), we might begin at the left limit of a vibration with the force at its greatest value and the velocity momentarily zero; then, as the time progresses (to the

right in the diagram), the velocity grows and the force diminishes, so that the velocity reaches its maximum a quarter of a period after the force, that is, the velocity is said to ***lag*** behind the force by this amount. Expressed as an angle (a complete vibration corresponding to 360°, as is usual) this lag is spoken of as a lagging phase angle of 90°.

This mechanical device (consisting of a force and an inertia) is closely analogous to an alternating E. M. F. [1] in a circuit with ***inductance.*** The inductance may be furnished by a part of a circuit with self-induction, such as by a coil. The effect of the inductance is to make the current in the circuit lag behind the E. M. F. just as the velocity lags behind the force in the curves of Fig. 26–2. We may replace velocity by current, and force by E. M. F. It often happens in alternating-current circuits that the current lags behind the E. M. F. by some angle between 0° and 90°. If inductance ***alone*** is present in the circuit, the lag is 90°. If resistance also is present, the lag becomes less than 90°, and the same thing can be shown to happen in the mechanical example considered above, if friction acts.

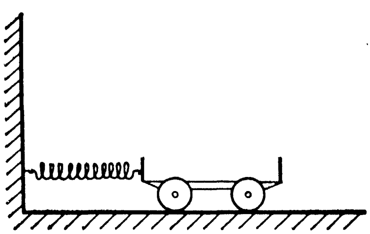


FIG. 26–3
A mechanical analogue of a capacity

Phase change due to capacity. In electrical lines there is always a certain amount of capacity, one line being near others, or near the surface of the earth, so that the system imitates the two plates of a condenser. This capacity is distributed all along the lines. Sometimes condensers are inserted in alternating current lines for purposes which are explained below (p. 419).

The effect of capacity is opposite to that of inductance; it advances the phase of an alternating current instead of retarding it. The reason for this will appear plainly by referring again to our mechanical model. We shall suppose now (Fig. 26–3) that the mass is removed, that the box itself is very light and that it is attached by a rather stiff but light spring to a rigid support. Now the hand forcing the spring to oscillate with simple periodic motion, as before, has to push with the greatest force to the right at

[1] The discussion is limited to oscillations which are of the type already called simple periodic motion.

the right-hand end of the swing, because the pull of the spring is then greatest while the maximum velocity to the right has ***preceded*** this by a quarter of a period. Thus the force and velocity are now represented by the curves in Fig. 26–4, with the velocity in the lead by 90°.

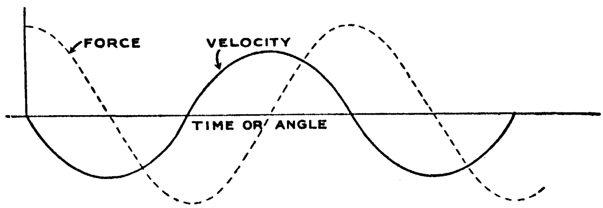


FIG. 26–4
Force and velocity in the mechanical model, or E. M. F. and current in a circuit containing capacity only

The alternating current in a circuit which contained nothing but capacity would act similarly, but the current would lead the E. M. F. by 90° in phase. A condenser acts in a manner analogous to a spring. When it is charged, it offers an opposing potential difference proportional to the charge on it, just as a spring resists with a force proportional to the displacement. The reciprocal of the capacity corresponds to the stiffness of the spring.

Impedance. If a circuit contains resistance and inductance, it obstructs the passage of an alternating current from both causes. The resistance uses up energy by converting it into heat. The inductance checks the flow by its opposing E. M. F., so that the current is diminished and less energy is expended. An interesting example of the effect of inductance is furnished by a "light dimmer," made of a coil of many turns which has a removable iron core (Fig. 26–5), connected in series with the A. C. lighting circuit. While the iron is being inserted, the inductance of the coil increases greatly, with the result that the resultant E. M. F. is reduced, and the light gradually fades.

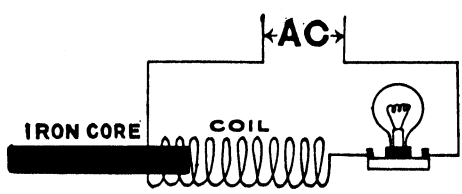


FIG. 26–5
A light dimmer

The combined effect of resistance, capacity, and inductance in a circuit is called the ***impedance.*** It can be found by a graphical

method. We have seen that an inductance by itself introduces a phase lag of 90° in the current. To find the impedance in the simple case when the capacity is negligible, we choose a reference line OA (Fig. 26–6) and draw a line OB at 90° to it, measured counter-clockwise. If the direction of OB represents the phase of the applied E. M. F., OA will then represent the phase of the current if inductance only is present. If resistance only were present, it would introduce no phase lag. We select a length OA just long enough to represent R, the resistance on some convenient scale. It can be shown that if we choose a length along OB equal to $2\pi fL$ on the same scale (where L is the inductance of the circuit in henries and f the frequency of the oscillations) then the diagonal of the rectangle formed by OA and OB, i.e., the line OC, represents in length the combined effect of resistance and inductance (that is, the impedance in this case), and the angle COA is the phase angle of lag between the current and the E. M. F. due to this impedance. Mathematically expressed, the impedance $= \sqrt{R^2 + (2\pi fL)^2}$, and the tangent of the phase angle is CA/OA, or $2\pi fL/R$. The quantity $2\pi fL$ is called the ***inductive reactance.***

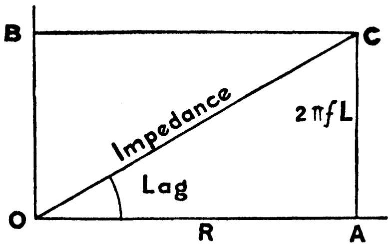


FIG. 26–6

Impedance of circuits containing resistance, inductance, and capacity. In case the circuit contains capacity (as well as inductance and resistance), its effect is to advance the phase of the current. The applied E. M. F. necessary to balance the effect of the capacity may be represented by a line OD, Fig. 26–7, drawn downward, at 90° to OA. It can be proved that the impedance introduced by the capacity is equal to

$$\frac{1}{2\pi fC},$$

where C is the capacity in *farads;* this quantity is known as the *capacity reactance.* The total impedance is now found by combining the three vectors, OA, representing the E. M. F. needed to balance the RI potential difference, always in phase with the current; OB, representing the E. M. F. to balance the effect of the inductive reactance, that is, an applied E. M. F.,

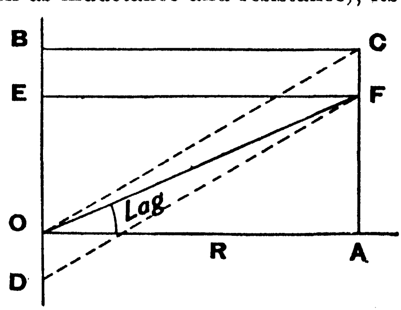


FIG. 26–7

whose value is $2\pi fLI$; and OD, representing the applied E.M.F. to balance the capacity reactance. Since OB and OD are opposite, it is simplest to subtract one from the other, thus obtaining OE, and then combine OE and OA to get OF, which represents the resultant applied E.M.F.). The angle AOF is the lag between the current and this resultant E.M.F. The effect of adding capacity is evidently to reduce the angle of lag, and this could be made zero if the capacity were of the right amount. The impedance in the most general case is given by the expression

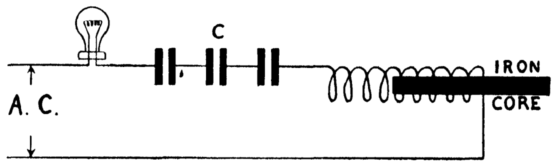


FIG. 26–8

$$\sqrt{R^2 + \left(2\pi fL - \frac{1}{2\pi fC}\right)^2},$$

but it would reduce simply to R in the case of zero lag, as can be seen from the manner in which Fig. 26–7 is constructed, since the only way in which the angle FOA can be made zero is to make F coincide with A. When the lag is zero we have a case of some practical interest, as the circuit is then a "resonant" one (p. 433). Evidently, if the impedance reduces to R, it is because

$$2\pi fL = \frac{1}{2\pi fC} \quad \text{or} \quad f = \frac{1}{2\pi\sqrt{LC}}.$$

Oscillations occur in such circuits and have a period

$$T = \frac{1}{f} = 2\pi\sqrt{LC}.$$

In power transmission a series resonant circuit is to be avoided because of difficulties which are likely to arise with it.

An interesting experiment can be shown illustrating the impedance of a resonant circuit. If an electromagnet with a removable iron core is available, together with a number of high-capacity telephone condensers, the inductance L and the capacity C can be connected in series in a circuit as shown in Fig. 26–8, with a lamp to indicate the current flowing in the circuit. When this arrangement is connected to the A.C. mains, and the inductance is gradually varied (by inserting the iron core in the coil) the lamp lights up brightly at resonance. The reason is that, as the formula shows, the impedance is then a minimum, being simply equal to the resistance.

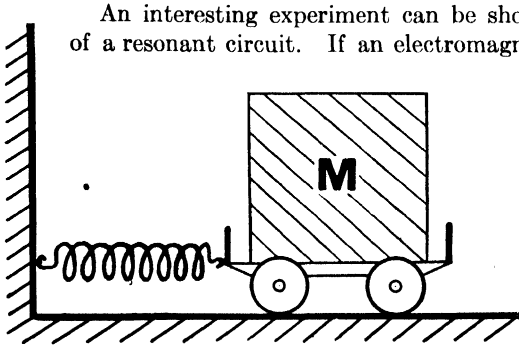


FIG. 26–9
Mechanical analogue for a resonant circuit containing inductance, capacity, and resistance

The properties of the circuit here considered are imitated by the mechanical model of Fig. 26–3, if a mass is put in the box to act like the inductance, while

the spring acts like the capacity as shown in Fig. 26–9. If this model is moved by hand with simple periodic motion, there is one particular period that can be chosen for the motion at which resonance occurs, and almost no force is required to move the model at this rate. Friction, acting like electrical resistance, is all that needs to be overcome, and there is now no lag between the force and the velocity.

Power in A.C. circuits. Power factor. In an ordinary circuit containing impedance, with an angle of lag between the E.M.F. (E) and the current (I), such curves as are shown in Fig. 26–10b

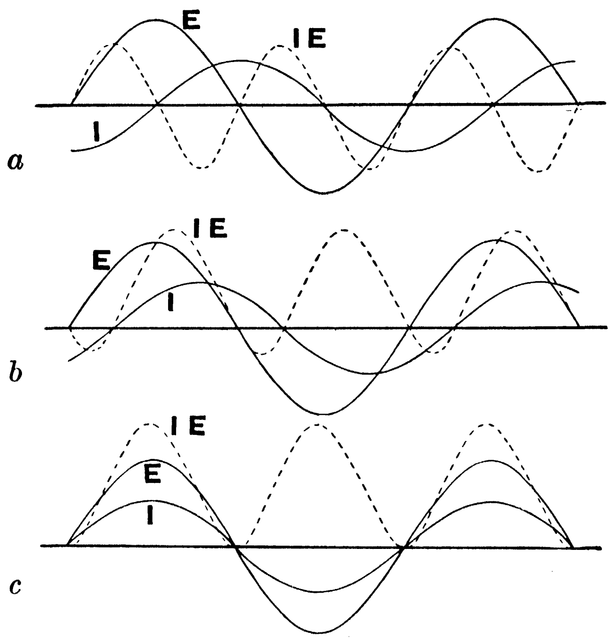


Fig. 26–10

might represent for a particular case the waves of current and E.M.F. in their proper phase relation to each other, the lag being shown here as about 45°. In this case it is interesting to calculate the power used in the circuit. To do this we must evidently multiply the value of I which exists at any instant by the value of E at the same moment, and average all such values over a complete period. In this way the curve IE is obtained, point by point. The power is positive whenever I and E are both positive, or both negative. When one is positive and the other negative, the product is negative, which means that during this time the power is

wire may be four times greater at a frequency of 200,000 cycles per second than it is for direct current.

This effect is due to induction. The changing current in the outer layers induces opposing E. M. F.'s in the inner parts of the same wire, so that with very high frequencies the current may even at times be running oppositely down the middle of the wire, in which case the resistance can be reduced by substituting a tube for a solid wire of the same size.

Alternating-current generators. Very large currents are generated in modern power plants at very high voltages, and the difficulty of securing good insulation in these machines becomes quite important. Largely on this account the rotating part of the machine (which was the armature in the generators previously

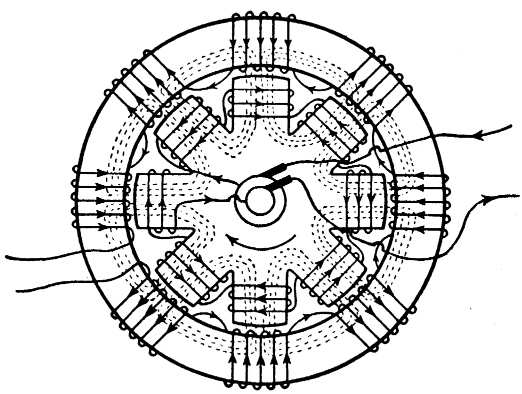


FIG. 26–11
The windings in a single-phase alternator

considered) is changed into a group of electromagnets, while the outer stationary framework now contains the coils of the armature. In addition, a small amount of direct current must be led through brushes and a split commutator into the revolving magnets in order to magnetize them. This direct current must come from a separate source, called the "exciter." Thus the large, induced alternating currents arising in the stationary armature coils are led directly into the external circuit.

Figure 26–11 shows a simplified sketch of the arrangement of the revolving magnets and the coils in generators of this sort, which are usually called ***alternators***. In this ("single-phase") arrangement the magnets by their revolution thrust magnetic lines into the coils and then withdraw them, all coils being affected

alike at the same time. The polarities of the magnets alternate around the rotor, and so do the windings of the coils, so that all coils furnish current in the same direction at each instant. As the coils are in series, a high voltage is easily secured. Figure 26–12 shows the two parts of an alternator, the ***rotor*** above, with the field magnets and a pair of rings by means of which

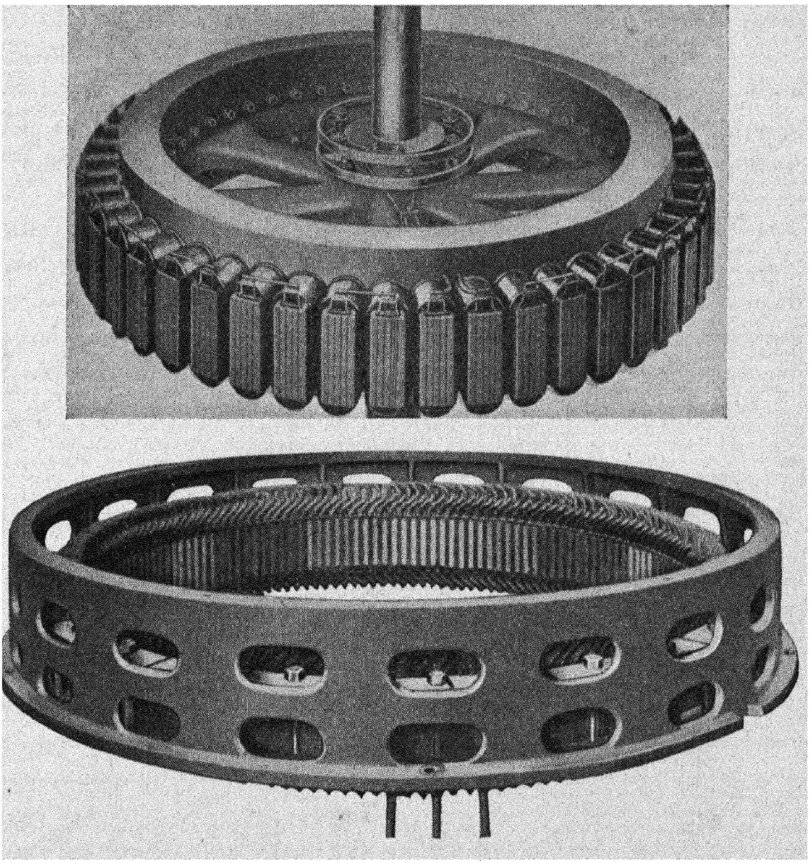


FIG. 26–12.
An alternator. (Courtesy of the General Electric Company)

the magnetizing current is brought in from the exciter, and the ***stator*** below, with its coils. This is a horizontal machine, such as is used with water turbines.

If the source of power is a steam turbine, a rotor with fewer poles is used, very sturdily built, so that it will not fly apart at the speed of the turbine; the high speed raises the rate of alternation to a suitable value, at the same time allowing the turbine to run

most efficiently. Figure 26–13 shows the revolving part (the "field") of an alternating current generator (3-phase) used with a steam turbine. It has two long poles, one of which is facing

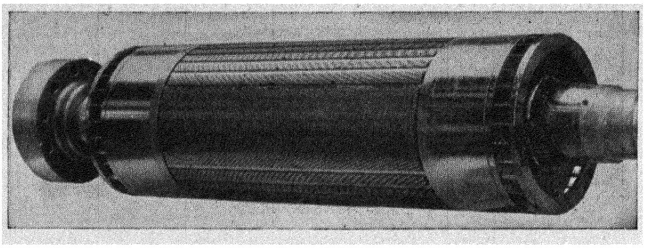


FIG. 26–13
(Courtesy of the General Electric Company)

the reader. The coils are sunk in slots and are wound around the poles. Figure 26–14 shows the stationary armature for such a machine.

The frequency is chosen usually as 25 cycles per second for power purposes and 60 for lighting (15 and 50 are common in

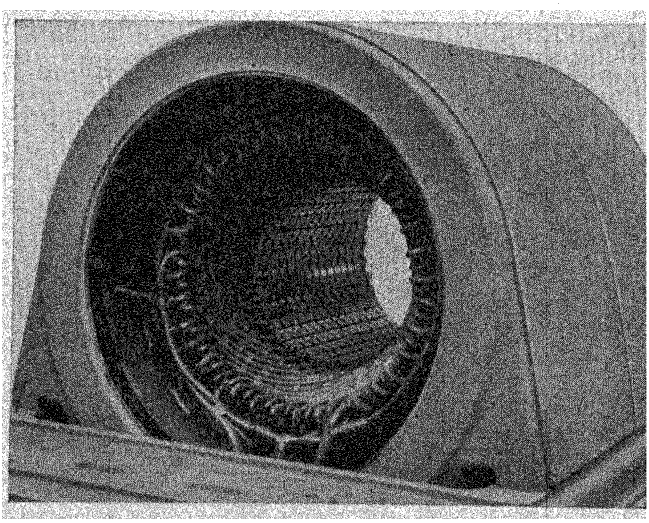


FIG. 26–14
(Courtesy of the General Electric Company)

Europe). The higher rate in a lighting circuit is advisable to prevent flickering of the lamps. The lower rate is preferred for power transmission because there is less energy lost at low frequencies in the magnetic circuits of the transformers, motors, etc., through which the current passes and for other reasons.

The voltages generated by large alternators range from about 2000 volts to over 13,000.

Transformers. Transformers are electrical devices by means of which alternating electromotive forces can be raised or lowered,

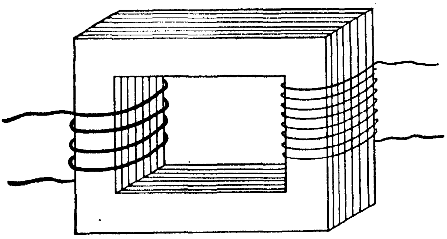


FIG. 26–15

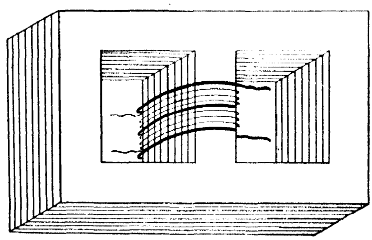


FIG. 26–16

and currents at the same time diminished or increased. We must consider how they do this, and why such changes are useful.

A transformer consists of a continuous silicon-steel core, built up of thin sheets, insulated by varnish or otherwise, making a magnetic circuit with practically no air gaps; around part of this is wound a primary and a secondary coil, usually one over the other, though sometimes on different parts of the core. The iron circuit may be a simple rectangular one, as in Fig. 26–15, or a form such as Fig. 26–16, in which case both coils embrace the central cross-piece, one wound over the other (shown as coarse and fine). A newer type, Fig. 26–17, called ***cruciform,*** consists of two such figures set in planes at right angles to each other, with the part, around which the two coils are wound, in common. A complete transformer is shown in Fig. 26–18, open and enclosed. One coil has more turns than the other. Either may be used as primary. If an alternating E. M. F. is connected to the one with the fewer turns, a higher voltage will be generated in the coil with more turns; and conversely. Thus the instrument may be used to "step up" the voltage or to "step it down."

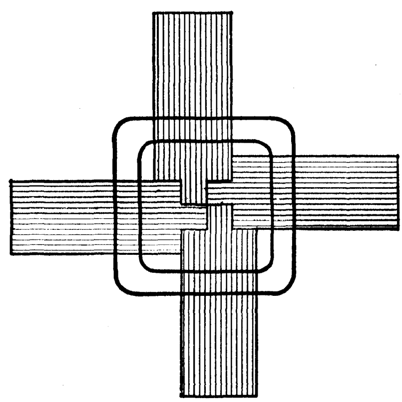
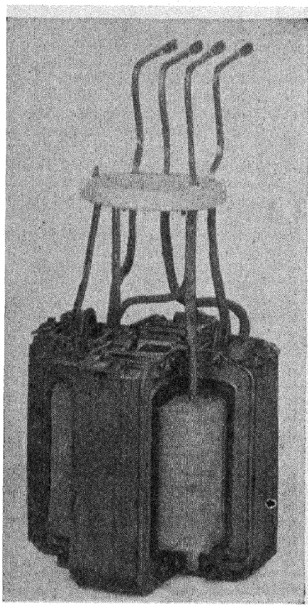
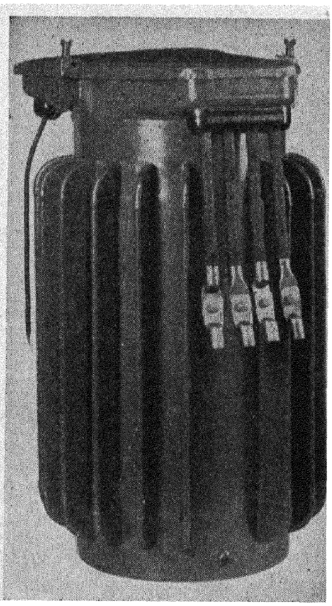


FIG. 26–17
The iron in a cruciform type. The coils are indicated by two loops.

Since these same lines "thread" the secondary and primary coils in each oscillation, it follows that the E.M.F. induced in that coil must bear the same ratio to the E.M.F. in the primary as the number of turns in the secondary bears to the number in the primary. Thus, if the primary has 50 turns and the secondary 2000, a primary E.M.F. of 1000 volts will be raised to 40,000 volts in the secondary coil.

If no current is being drawn from a transformer, the self-induction of the primary coil is so great that the back E.M.F.

a FIG. 26-18 *b*

(Courtesy of the General Electric Company)

reduces the current in the primary to a very small amount. In other words, the current lags almost 90° in phase behind the impressed voltage, and no power is required except the very small amount needed to compensate for the losses in the iron. When current is drawn from the secondary, the phase angle of lag is reduced, and more power is then drawn from the source.

Step-down transformer. Electrical welding. If a transformer has many turns in the primary and only one or two (of *very* heavy copper) in the secondary, so much current will flow through the secondary that two iron rods in contact with each other as in Fig.

26–19 at *C*, which form part of this secondary circuit, may be welded together. The heat is generated where the resistance is, i.e., at the point of contact. This experiment illustrates a method in common use, as, for instance, in welding street-car rails together.

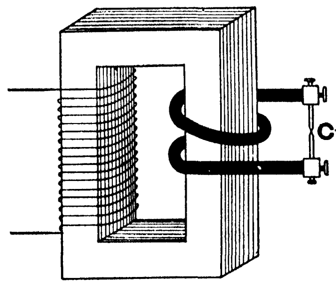


FIG. 26–19

Power loss in transformers. If the iron core of the transformer is carried through a complete magnetic cycle by one cycle of the current, it has followed through a complete hysteresis loop (p. 381) whose area is proportional to the loss of energy which has occurred. In the best silicon steel which is used in transformers this loss is only 2 to 5% of the energy supplied by the source. Hence if I and e are the current and the E.M.F. in the primary coil, and i and E those in the secondary, we have (within this limit of accuracy) $Ie = iE$, or the current is as much reduced as the voltage is raised.

Power transmission. In alternating, as in direct, currents the loss of power in transmitting a current of I amperes through a resistance of R ohms is RI^2 watts. If electrical power has to be transmitted over a long distance, this loss may become serious. For instance, if an attempt were made to transmit 100 kilowatts (100,000 watts) at 500 volts over a line of 1 ohm resistance, the current would be 100,000/500, or 200 amperes, and the heat loss in the resistance would be $1 \times (200)^2$ or 40,000 watts, or 40 kw. This is, of course, 40% of the power transmitted. If the voltage could, however, be raised to 50,000 volts, the current would be reduced to 2 amperes, and the loss on the line would be 4 watts, or 0.004%, which is negligible.

Thus it happens that in large power stations (e.g., at waterfalls) situated at some distance from the place where the power is to be used, the generators are often made to deliver power at over 10,000 volts to transformers, which may raise the voltage to 200,000 or more for transmission over the long line. If insulation difficulties can be overcome, and the electrical troubles due to the effects of lightning reduced, higher and higher voltages may come to be used for power transmission. At present the limit for power transmission is about 220,000 volts. For experimental purposes over 5,000,000 volts are being generated.

Such high voltages are extremely dangerous on account of the ease with which they can jump an air gap or cause a glow-discharge out into the air. Figure 26–20 shows the "corona" glow from a conductor of 4.5 cm. diameter subjected to 800,000 volts in air. A discharge is seen passing off into the air at many points along the conductor. For distribution of power in cities, it is customary to reduce the E.M.F. to about 2300 volts at the outskirts of the city; then at each house (or group of houses) a small transformer brings it down to 110–220 volts for domestic use.

FIG. 26–20

Corona discharges. (Courtesy of the General Electric Company)

Rotary converters. It is often desirable to change from A.C. to D.C. One way of accomplishing this is by the use of a *motor-generator* set; thus alternating current may be used to run an A.C. motor, and this may drive a D.C. generator mounted on the same shaft. A neater and more efficient way of doing it is to have a machine, called a *rotary converter* furnished with a rotating armature with commutator and *also* with slip rings. In this case an alternating current may be sent through the slip rings into the armature coils, and the machine then acts like an A.C. motor, except that it has nothing but itself to run. The alternating current also reaches the commutator, and brushes bearing against this can take it off as direct current just as they do in any direct current generator.

The change from D.C. to A.C. may be accomplished in like manner but is not so often required.

Rotary fields. Two-phase currents. An alternating current can create a magnetic field which rotates without requiring the rotation of any material. This curious effect may be produced by what is called two-phase current, which consists of two currents, flowing in separate circuits, one of which is permanently a quarter period (90°) ahead of the other in phase. Such currents can be

taken from different coils in the same generator. Thus, if coils A and B (Fig. 26–21) are part of a 4-pole A.C. armature and the magnet NS (which is, of course, really an electromagnet) is revolved, the currents generated in coils A and B are 90° apart in phase. If now these currents are led (by four wires) to similar coils P and Q in another machine (A to P and B to Q), the magnetic field in the second machine will rotate. This is easily seen if we note that at one moment the coils P are fully magnetized, say in the direction from top to bottom; a little later this field has died out and the coils Q have become magnetized instead, so that the field lines now run from right to left; later they go from bottom to top, and still later from left to right. At intermediate

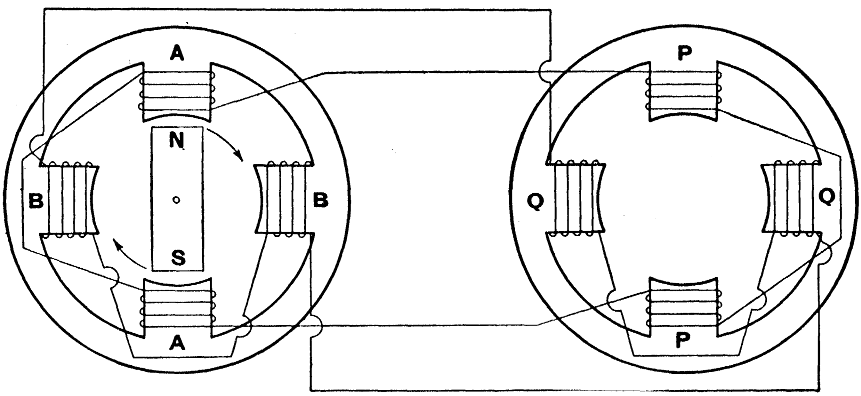


FIG. 26–21

times it is easy to show that the combination of the two fields gives a resultant in an intermediate position; so that in the central space there exists at all times a steadily rotating magnetic field of practically constant strength.

Induction motor. A sort of "squirrel cage" is made of heavy copper bars whose ends are welded to two circular conducting rings. The bars are imbedded in a laminated iron core, so that only their projecting ends are visible. This device, when suitably mounted on a shaft, becomes the rotor of an induction motor, as shown in Fig. 26–22. If this rotor is mounted in a magnetic field which is rotating in the way described in the last paragraph about the same axis, the magnetic lines will cut the copper wires, and there will be large currents induced in them. By Lenz's law, the magnetic field produced by these induced currents will be in such a direction

as to oppose the change that is occurring; that is, they will drag the copper bars after the field, tending to make them go in the same direction and at the same speed. The wires will never catch up with the field, but will slip back by an amount which depends on the work being done by the rotating device (rotor). If the rotor is doing heavy work, it runs more slowly, the magnetic lines pass by it quickly, large currents are induced and thus large forces are produced to drive it.

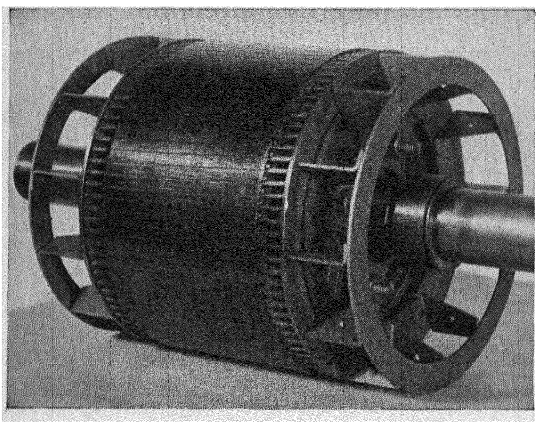


FIG. 26–22
The rotor of an induction motor. (Courtesy of the Wagner Electric Corporation)

Single-phase motors. Two-phase alternating current is not commonly used. The simple sort we have been discussing earlier is called single-phase alternating current. Motors capable of running on this sort of current may be of two sorts.

The *synchronous motor* is a machine built like a single-phase generator. If it is started by some separate source of power before it is connected to the alternating source which is to run it, and reaches such a speed as to make it precisely synchronous with that current (i.e., the currents in its coils alternate at the same rate and in the same phase) then the circuit may safely be connected, and current will flow through each armature coil at precisely the right time to attract the corresponding magnet on the rotor. If the "load" on the machine is increased, the rotor lags through a small angle, the forward pull increases and the necessary power is thus supplied for the increased load without any slowing down on the part of the motor.

The *single-phase induction motor* is similar in that it needs help in getting started. Though it seems very different from the two-phase machine, it acts in essentially the same way.[1]

Three-phase currents. If a generator is devised so that successive coils, or groups of coils, in its armature have currents generated in them which differ in phase by 120° (as indicated in

[1] The reader interested in details will find a full description of alternating instruments and machines in "Electrical Engineering," Vol. II, by C. L. Dawes, second edition, 1928, (McGraw-Hill), and in "Electrical Circuits and Machinery," Vol. II, by J. H. Morecroft and F. W. Hehre, 1924, (John Wiley and Sons).

a simplified way in Fig. 26–23), then three-phase current may be taken from these coils. One would suppose that six wires would be necessary to conduct away the three currents thus generated, but three wires are sufficient as they may be joined together in pairs. The current of one phase may be regarded as going out at any moment on one wire and returning on the other two (or going out on two and returning on one). Thus three-phase current requires fewer transmission wires than two-phase current does, and it is very much used in the distribution of electrical power. The advantage which it possesses over single-phase (or simple alternating) current is that larger currents can be carried for the same weight of wire.

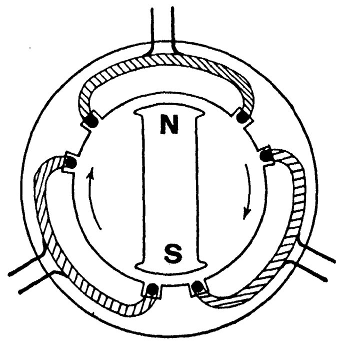


FIG. 26–23

The ***three-phase induction motor*** is very commonly used. It requires a three-phase current which produces a rotating magnetic field in which the rotor moves. The machine acts in much the same manner as the two-phase induction motor already described.

Motors for A. C. or D. C. Some motors are marked "for A. C. or D. C." These are series-wound motors in which a reversal of the current reverses the polarity of both the field coils and the armature coils, leaving the direction of the force between them the same as before. They will thus operate successfully on single-phase alternating current as well as on direct current.

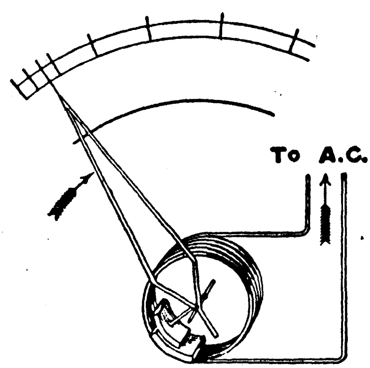


FIG. 26–24
A Weston iron-vane instrument

A. C. ammeters, voltmeters and wattmeters. For measuring alternating currents the ordinary galvanometer is unsuitable, since the force tending to turn the coil when the current is flowing in one direction reverses itself so frequently that the inertia of the coil prevents any motion from occurring at all. The following types are, however, suitable for either A. C. or D. C. measurements.

The *hot-wire ammeter* (p. 375) is a device which has already been described. Only the thermoelectric forms of this instrument are good for very precise measurements.

The *plunger type* of instrument has a piece of soft iron which is attracted

into or toward the coil through which the alternating current is flowing. This attraction occurs equally well whichever direction the current takes, since the magnetism of the soft iron reverses with the polarity of the coil attracting it. The motion of the plunger makes an indicating pointer travel over a graduated scale. A modification of this which is found in a very useful type of *iron-vane* instrument is shown in Fig. 26–24. Here a piece of soft iron is fastened inside a coil. A second piece is mounted on a framework capable of rotation. The alternating current in the coil magnetizes both pieces always in the same direction and on account of their shape and relative position they always repel each other and their mutual repulsion moves the one which is free. A spring opposes this motion and a pointer indicates its amount on a graduated scale.

The *dynamometer type* of instrument has two coils, one fixed and the other movable; the fixed coil is usually in two parts, one on each side of the movable coil. The current flows through both in series and reverses at the same time

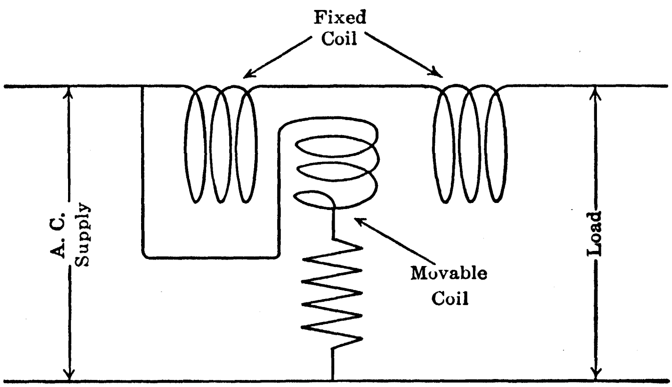


FIG. 26–25

in both. If they are so wound that the movable one turns toward the other when the current is going in one direction it will do the same when the current is reversed. Hence a steady deflection is secured with an alternating current. As it is difficult to arrange a movable coil to carry a large current, these instruments are not used in the form of ammeters; but voltmeters made on this plan are quite satisfactory.

The *wattmeter* is a form of instrument of the dynamometer type in which the fixed coil, in two parts, is connected in series with the current, and the movable coil is across the mains with an added high resistance in series. Figure 26–25 shows the circuit arrangement and Fig. 26–26 one of the instruments. The current in the series coil builds up a magnetic field proportional to I. The current in the movable coil is proportional to E, as it is in any voltmeter. The deflection is proportional to the product IE, or the number of watts being measured. If the movable coil is not held by springs, as it is in all portable instruments like ammeters, etc., but is mounted to rotate like a motor, the number of its rotations in unit time is proportional to the energy used in that time, provided that proper friction is supplied. If this motor is so connected as to move pointers over graduated dials, the instrument becomes

a *watt-hour meter*, which records *power-time*, or energy. The friction is ingeniously introduced by eddy currents induced in a rotating disc by a permanent magnet, and is proportional in amount to the speed of the disc.

Following Faraday, Barlow discovered that a current traveling in a circular disc from its center to its rim, and leaving by way of a dish of mercury, into which the disc dips, would make the disc rotate if there were a magnetic field perpendicular to its plane. This makes a so-called *unipolar motor*. This principle is used in another successful type of *watt-hour meter*. The field coils take the place of the movable coil in the type just described, that is, they produce a field proportional to the voltage, while the current in the rotating disc is in series. Hence the rate of rotation is proportional to IE. Friction

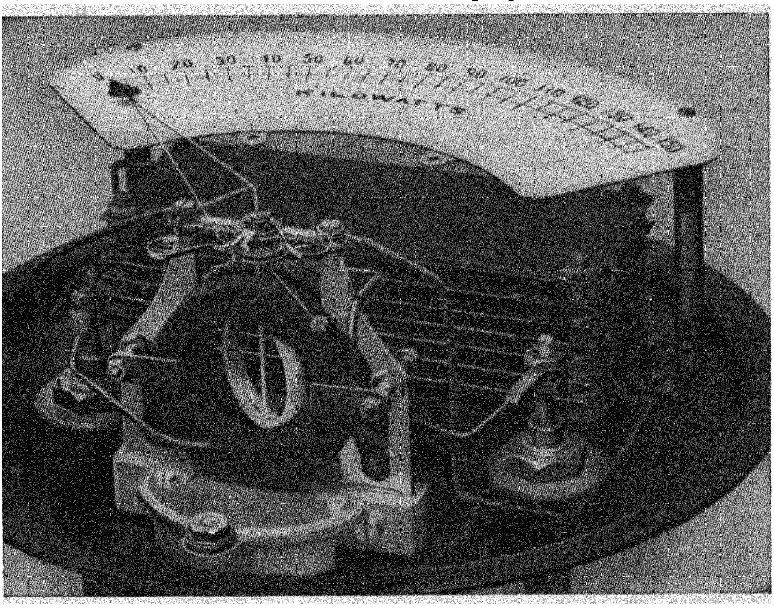


FIG. 26–26

A wattmeter. (Courtesy of the Weston Electrical Instrument Company)

is provided in the same manner as before, proportional in amount to the speed of the disc. In some of the best of these instruments the disc actually floats in mercury; the current flows through the disc rather than the mercury because the disc is made of copper which is a much better conductor.

PROBLEMS

1. A transformer for house lighting purposes is perhaps two feet high. Why cannot much smaller ones be used, the ratio of the number of turns in the two coils being kept the same?

2. A transformer has 100 turns in its primary coil and 5000 in its secondary. If a 100-volt line sends 10 amperes through the primary, find the voltage, current, and power in the secondary circuit, assuming no losses in the transformer.

3. A step-down transformer has 2000 turns in its primary coil and 40 in its secondary. The primary is connected to a 2000-volt line, and the secondary delivers 50 amperes. Find the secondary voltage, primary current, and power transmitted, assuming no losses in the transformer.

4. An alternator with 16 poles is producing 25-cycle current. Each pole has 500,000 lines, and each armature coil has 20 turns. How many lines are cut by each armature coil when one pole goes by? How long does this operation take? What is the average E. M. F. generated in each armature coil during this time? What is the average E. M. F. in the whole 16 coils connected in series? (N.B. This is not the "A. C. voltage" of the machine. Why not?)

5. A 60-cycle 110-volt circuit has 12 ohms resistance in series with 0.2 henry inductance and 24 microfarads capacity. What is the current in the circuit?

6. An ammeter shows that an A. C. generator is delivering 21 amperes, and the voltmeter reads 220 volts. A wattmeter shows that 4 kw. are being delivered. What is the power factor?

7. A single-phase induction motor takes 25 amperes, and the current lags somewhat behind the E. M. F. (which is 220 volts) so that the power factor is 0.87. How much power does the motor take?

CHAPTER 27

ELECTRIC OSCILLATIONS AND WAVES

Resonant circuits. If a circuit is set up, as in Fig. 27–1, consisting of a condenser C and an inductance L, and any sort of an electrical disturbance is made in it, say by momentarily charging the condenser, there will arise a series of electrical oscillations in the circuit which are usually extremely rapid, and which soon die out unless maintained by some outside source. The formula giving their rate of vibration has already been derived (p. 417). Such a system is like a spring supporting a heavy mass, as in Fig. 27–2. The yielding of the spring at the extremity of the vibration corresponds to the charging of the condenser; the electrical energy is then in the electric field between the plates. The velocity of the mass in the middle of the swing corresponds to the current through the inductance, and then the energy is largely in the magnetic field there. If there is a little friction in the mechanical case the oscillations will slowly decay. In the electrical case the inevitable resistance in the circuit produces a similar result. But there is another way also in which energy is dissipated. In the mechanical case, if the vibrations of the spring were sufficiently rapid, we should hear a sound coming from it; which proves that

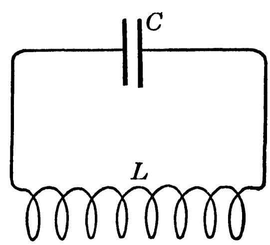


FIG. 27–1

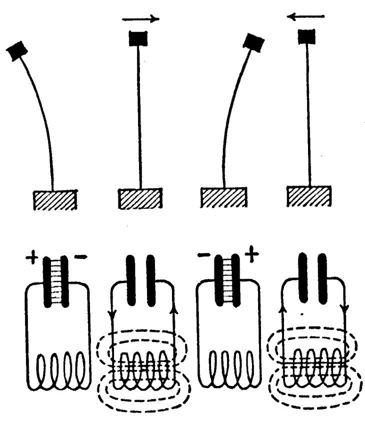


FIG. 27–2

it gives rise to air waves by its motions, and these must naturally carry off some of its energy. In the electrical case there must be something similar, the disturbance somehow creating waves of an electric nature.

Electromagnetic waves. A charged body has an electrical field about it which extends to an indefinite distance, though we may not be aware of it beyond a few inches or feet from the body. This field is more noticeable the farther apart are the opposite charges which are involved. Thus a condenser has comparatively little in the way of "stray" field lines, most of them crossing directly from the positive charge on one plate to the negative on the opposite. A metallic ball or a vertical wire if charged, say positively, has a field which spreads out, perhaps through quite a long distance, if the conducting body is more or less by itself, the lines ending on negative electricity, as usual. This negative is induced on conductors in the neighborhood, if there are not already negative charges near by.

If the charge on an isolated ball is suddenly removed, the field about it must disappear. An interesting question arises as to whether it all disappears instantaneously, or whether what we might call a wave or pulse of annihilation starts from the ball and travels out, destroying the field as it goes. An act propagated instantaneously through extended space calls for super-physical agencies. We turn, therefore, to the second alternative as the more reasonable. Now suppose that the isolated ball is charged alternately positively, negatively, positively again, and so forth. There must then travel out from the ball an alternating something, an action first establishing a positive field, then destroying it, then establishing a negative field, which is in turn destroyed, and so on. This sort of action spreading out from a source at which charges are oscillating is called an ***electromagnetic wave.*** The best description that can be given of it is none too good; it must be objectionably vague. We have no idea as to the nature of the disturbance in space which we call an electric field. Changes in it were called "electric displacements" by Maxwell, and this term is convenient even in cases where there are no electrons, nor anything else that we know of to be displaced. Making use of it, nevertheless, we may say that any electric displacement, like a moving charge, creates a magnetic field at right angles to the motion; hence we cannot have a simple electric wave without the accompaniment of magnetic

effects. Thus the term "electromagnetic" arises, rather than simply "electric" in referring to the wave.

These waves, as we shall see later (p. 555) are the ones we use in radio communication. They travel with the speed of light and are like waves of light, heat radiation, and X-rays in their nature, though not in their wave-lengths.

Waves from oscillating circuits. A circuit such as that shown in Fig. 27–1 has a small external effect when it oscillates, but it does, nevertheless, radiate out electromagnetic waves which carry away some of the energy in the circuit, thereby producing what is appropriately called a radiation loss. This is analogous to the energy radiated as sound from a vibrating tuning fork, but, as in sound sources, the circuit which radiates the best is one with a large and effective connection with the outer medium, that is, one with a large external field. Thus, a tall vertical conducting wire connected through a coil to the earth, like the antenna often used in radio circuits, has a large external field when charged, and unless continually supplied with energy loses it so rapidly by radiation that only a few oscillations occur before it is all gone.

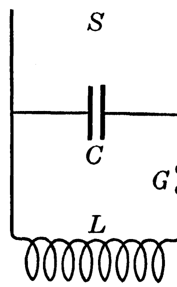


FIG. 27–3

Sources of electric oscillations. Sparks. We must now consider means of making electric circuits oscillate. One of the easiest methods is by means of sparks. In Fig. 27–3, S is a source of high-potential electricity (e.g., the secondary of an induction coil) connected to a large condenser such as a Leyden jar, C, which is part of a circuit containing an inductance L and a spark gap G. The source charges the condenser which then causes a spark at G, and oscillations occur between C and L through the ionized vapor in the spark gap. One might suppose that the charge on C would be neutralized without any oscillations, but the inertia of the moving charge (i.e., the inductance in L) prevents this, just as the inertia of a pendulum causes it to go past its middle point and oscillate back and forth until its energy is eventually all dissipated by friction. Here the spark offers some resistance, and this, together with the radiation loss and the other resistances in the circuit, reduces the number of oscillations to perhaps 50 or 100 if the inductance and capacity are large and the spark short, or to 2 or 3 if the spark is a long one and the coil is small. Sparks, such as the long, thin ones (several inches

in length) obtained with large induction coils involve hardly more than one oscillation, so rapidly are they "damped." If an image of an oscillating spark is cast by a good lens on a moving photographic plate, a picture like Fig. 27–4 is obtained. Here one sees a practically instantaneous discharge at first, followed by a set of

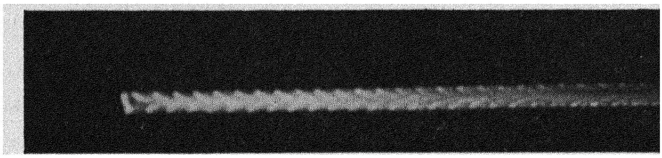


FIG. 27–4
The oscillations in a spark, photographed on a moving plate

discharges emanating first from one pole and then from the opposite, to the number of thirty or so. The rate of oscillation in such a case as this can be measured if the speed of the plate is known, and it proves to be very high, often several millions per second.

A method of producing continuous oscillations will be considered below (p. 459).

Electric waves along wires. An interesting experiment can be performed which demonstrates the existence of electric waves along wires. A source S, such as an induction coil, is connected by two wires A and B (Fig. 27–5) to two semicircular conductors, the connection on the side B being through a small spark gap in oil. Below these lies a loop connected to two parallel wires which in the form here shown are enclosed by a tube containing air at low pressure. The spark at the connection B produces oscillations in the two semicircular pieces, which in turn induce oscillations in the loop and in the long wires. The waves traveling down the wires are reflected at their ends and may set up a series of standing waves which show themselves by a wave-like distribution of corona discharges in the tube, visible in a darkened room. These give a vivid picture of the wave system and make a direct measurement of the wave-length possible.

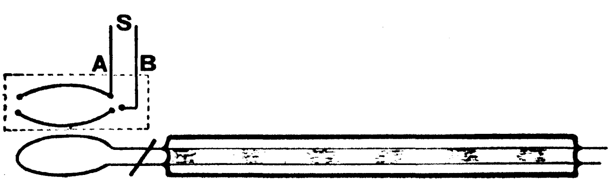


FIG. 27–5

The experiment shows that waves exist in the space around the

wires as well as in the wires themselves. An alternating current is thus not a thing confined to the conductor "in" which it flows, but it has an external field, whose effects are often evident. In fact, if Faraday's point of view is correct, the most important features of a current lie in the field and not in the wire carrying it. Instead of thinking of a current as a rush of electrons in a wire, we might regard it as a flow of lines in the field, each line ending at any instant on a charge in the wire.

It is interesting to note that waves along parallel wires, as in the experiment just described, have been shown to move with the speed of light. The standing waves can be used to find the speed, just as was done in the case of sound (p. 256). The wave-length is measured directly, the frequency is determined from the formula

$$\frac{1}{f} = 2\pi\sqrt{LC},$$

where L and C are found from other experiments, and the velocity is then obtained from the equation $V = f\lambda$. The simplest picture of the passage of such waves along a wire is to imagine the electrons rushing through the wire with the speed of light, carrying the ends of the electric lines of force with them; but this cannot be correct. There are good reasons for believing that an electron can never be made to move as fast as light, without giving it an infinite amount of energy; and, besides, the space inside a solid conductor would not appear to be a good track along which exceptional speed records could be made by an electron. It is too full of obstacles. There are other ways in which a wave could be generated. We are familiar with waves which we make in a rope by shaking one end, and these may travel at a considerable speed while the particles of the rope do not move in the direction of the waves at all, but transversely to it. Thus electrons might be imagined to move in and out through a very small distance at right angles to the wire, charging the surface of the metal negatively when they are on the outside and positively when they retire within. Very small motions of the electrons in proper phases would generate a wave at right angles to these motions. But we know as yet too little about the nature of matter to be able to say that this is what really happens; we can say only that the waves in air travel with the speed of light, while the electrons move more slowly, and probably through short distances only.

Tuned circuits. When oscillations are occurring in a circuit and sending out waves into the surrounding region, any other circuit near by ***may*** pick up these oscillations by resonance (p. 272), provided its natural rate of oscillation is the same as that of the incoming waves. It is shown on page 417 that the time of a complete oscillation of a circuit is $T = 2\pi\sqrt{LC}$, where L is the inductance (in henries) and C the capacity (in farads). The formula reminds one of

that of a pendulum (p. 233). Changes in either the inductance L or the capacity C will enable us to "tune" one circuit to another, and then some detecting device may make it possible to receive signals in the tuned circuit. This is what one does in a radio receiving apparatus by turning a dial, the adjustment usually made being in a variable air condenser, as in Fig. 27–6. It has many plates, in two sets. Each plate of one set may be inserted to a greater or less extent between the others, without touching. This motion alters the capacity of the condenser continuously. Naturally such a condenser will not serve for any use involving high potentials unless immersed in oil, as sparks readily jump across the air separating the plates.

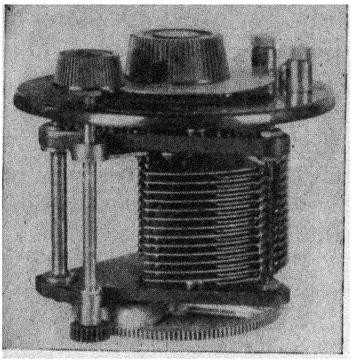


FIG. 27–6
A variable air condenser

Tuning may also be accomplished by turning one part of a coil with reference to the rest, thus continuously altering the inductance of the combination.

Underground waves. Electric waves travel fairly well through the earth. Distant radio stations can be heard in deep mines. Waves started near the surface may spread through the ground and be affected by the presence of *bodies of ore*, or even by a gas pipe, so that the hidden metallic bodies may be detected from above. A cable lying on the harbor bottom may send out waves through the water which can be heard by two "ears," consisting of vertical coils on each side of a ship, connected to telephones. When the sounds are equally loud in each ear, the ship is over the cable, and it may thus be guided through a tortuous channel.

PROBLEMS

1. A resonant circuit can be set up to have a period of two seconds, like a "seconds pendulum." If a capacity of 1000 microfarads (0.001 farad) is available, how large an inductance would be required?

2. The velocity of electric waves (and light) being known to be 3×10^{10} cm./sec., what is the frequency of an electrical oscillation if the glowing patches in the tube of page 436 are 80 cm. apart? (N.B. As in sound, adjacent antinodes are not one wave-length apart.)

3. Find the wave-length given by a spark in which 40 oscillations occur in a total time of 0.001 second.

CHAPTER 28

CONDUCTION OF ELECTRICITY THROUGH GASES

Conduction through a tube, 439; cathode rays, 440; deflection of cathode rays by magnetic and electric fields, 441; velocity of cathode rays, 441; the ratio of charge to mass, 443; the charge borne by the electron, 444; the mass of the electron, 445; variation of mass with velocity, 445; discharges at moderately low pressures, 446; positive rays, 446; the mass spectrograph; isotopes, 447.

Conduction through a tube. A long glass tube, as in Fig. 28–1, fitted with electrodes consisting of small plates sealed in at the ends and connected to an outside circuit, has a side tube at P leading to a good vacuum pump so that an electric discharge can be sent through the tube at any pressure. G is a spark gap of, say, 5 cm. length, connected as shown. If A and B are connected to a good induction coil, and the air in the tube is at atmospheric pressure, a spark will pass at G, but nothing will occur inside the tube. As the pressure in the tube falls to about 10 mm., thin, irregular streaks of bluish light flash through the tube, and the spark at G then ceases. It is now easier for the discharge to pass through a long column of low-pressure air than through the much shorter gap G outside. At about 1 to 3 mm. pressure the discharge fills the tube with a reddish glow, known as the positive column, except that at the negative terminal the light remains bluish and there is a gap between the two colors. Then, if the pump is a good one, and the apparatus free from moisture, the glow becomes fainter and whitish, the dark gap between the two colors disappears as they fade, and close to the cathode (negative terminal) there appears a new dark space surrounded by a faint glow (Fig. 28–2). As the exhaustion proceeds, the cathode dark space increases, the glow in the middle and on the positive side of the tube (the "positive column") fades away and at a pressure of

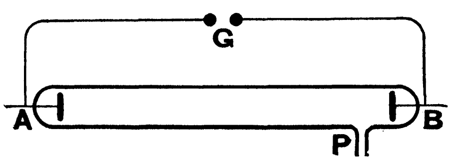


FIG. 28–1

about 0.02 mm., when the dark space next the cathode has expanded so as to occupy a large part of the tube, a new effect is seen. A glow (usually greenish) appears in the glass walls themselves, brightest near the negative terminal, the cathode. If the tube is then so treated as to produce a still higher degree of exhaustion within it, the discharge ceases altogether. This "high" vacuum is beyond the powers of an ordinary pump and cannot

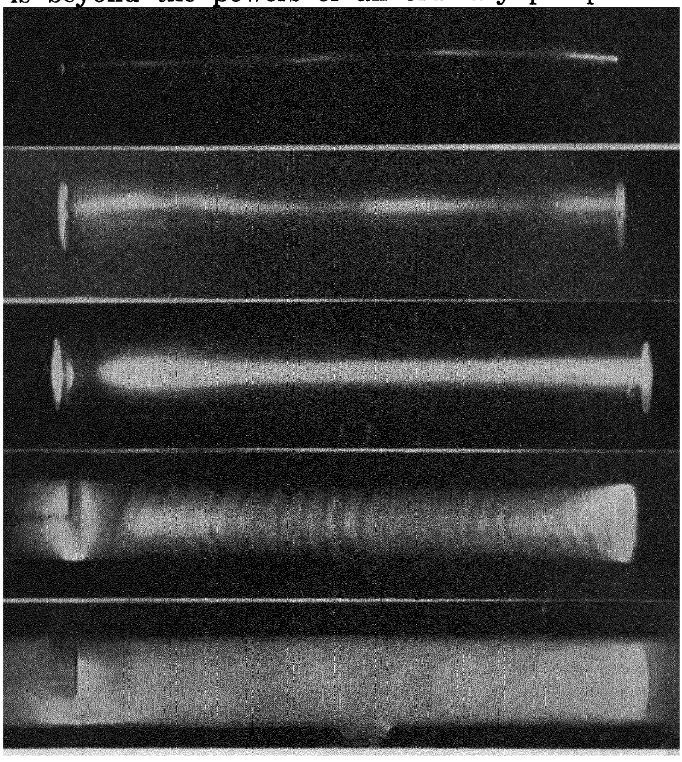


FIG. 28–2
The discharge in a tube at diminishing pressures

usually be reached without giving the tube a prolonged heating to bake out occluded gases.

This description does not apply to the case in which one of the electrodes is a hot wire; other phenomena then occur (p. 452).

Cathode rays. The glass glowing in the neighborhood of the cathode, as in the last paragraph, can be shown to be under the bombardment of something proceeding from the cathode in straight lines. This is most easily done in a tube of the form of Fig. 28–3, in which a figure cut out of sheet metal is mounted out

in front of the cathode (C) with a more or less flat glass wall W behind it. The anode (+) is placed in a branch at one side; its exact location is unimportant, as the rays from the cathode seem to pay no attention to it.

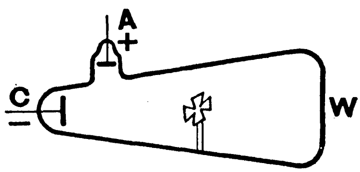


FIG. 28–3
A cathode-ray tube

In this tube, when exhausted to the proper degree and connected to an induction coil, there appears on the wall W a shadow of the metallic figure, which is not found if the connections are reversed. This shows that the "cathode rays" really do proceed from the cathode in straight lines, and that they are stopped by a piece of metal. If the cathode plate C is concave, the rays can be made to converge so as to form a small concentrated spot, at which a great deal of heat may be seen to be developed, if a piece of thin metal is placed there to stop them. Such experiments suggest that these rays consist of small bullets shot out of the cathode. J. J. Thomson, who discovered their nature in 1897, called these bodies "corpuscles." We now know them as ***electrons,*** detached from atoms and here flying about by themselves. The discovery of these facts was one of the greatest single forward steps in the history of physical science. The method by which it was made will be considered in the paragraphs below.

Deflection of cathode rays by magnetic and electric fields. An important experiment leading to this discovery came from exposing the cathode rays to a magnetic field. If we repeat the shadow experiment (Fig. 28–3) and bring a strong magnet near the stream of rays leaving the cathode, we can easily deflect them so that the shadow which they cause is seen to move. The direction of the deflection is at right angles to the magnetic lines and proves that the cathode rays are equivalent to a stream of negative electricity leaving the cathode. We should not, of course, expect anything but negative electricity to be driven ***away*** from the cathode. We come to the same conclusion if we arrange matters so that the cathode rays have to shoot through the space between the plates of a charged condenser. Here the electric field deflects the rays toward the positive plate.

Velocity of cathode rays. The characteristics of the cathode ray particles were settled by a series of quantitative experiments devised by J. J. Thomson, which we must now consider. In the

first of these, a tube contained a small cathode C, Fig. 28–4, and an anode A in a side branch. There were two metal screens H with a small hole in each which limited the beam of cathode rays to a narrow streak. This fine ray passed through an electric field between charged condenser plates which, if charged as shown in Fig. 28–4, would deflect the stream of particles ***upwards.*** In this same space external coils (not shown in the figure) produced a magnetic field inside the dotted circle perpendicular to the plane of the figure and directed away from the reader. If the stream of cathode rays consists of a mass of negative particles moving to the right in the figure, this is equivalent to a current of positive electricity moving from right to left. The magnetic field produced by such a current would generate lines going into the paper above the line of current, and hence crowding the magnetic field lines on this side. This would drive the "current" ***downwards,*** if it were flowing in a wire, and presumably also if it were flowing as a stream of charged particles in space. Hence the magnetic field produces an effect in a direction opposite to that produced by the electric field, and these two can be balanced against each other, so that the stream of cathode rays goes through undeflected. A screen S, covered with a material such as zinc sulphide, which glows brightly ("fluoresces") when the rays strike it, enables the experimenter to observe when the effects of the magnetic and electric fields exactly balance. Means of measuring the intensities of each sort of field are well known. The magnetic field is confined approximately in the same space as the electric field; each does in practice "stray" somewhat, but the effect of this can be allowed for by calculation.

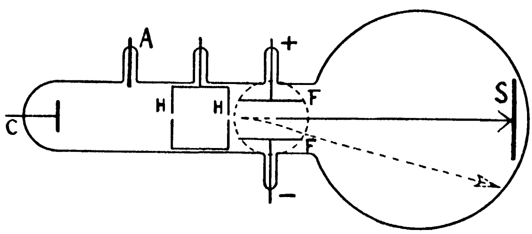


FIG. 28–4

A particle having a charge e will be acted on in an electric field of intensity E by a force of eE dynes (E–M units must be used here). The same particle having a velocity v and moving perpendicularly through a magnetic field of intensity H will be acted on by a force which we can find by the following reasoning. We know from page 390 that a wire of length L perpendicular to a magnetic field and carrying a current I (E–M units) is acted on by a

force equal to HLI dynes. This is the same as HLQ/t, if Q is the quantity of electricity passing a given point in time t, and this can also be written as HQL/t, or HQv, if $v = L/t$. The same force acts on a current I in a length L of wire as on a charge Q moving with velocity v. If we replace Q by e, the charge on our particle, the force becomes

$$F = Hev.$$

If the magnetic and electric forces are equal, so that the stream of particles passes through undeviated, it follows that $eE = Hev$, or $v = E/H$. This remarkable result gives us the velocity of the particles directly, before we know anything of their nature. We find that cathode rays move in these experiments with a speed not far from one-tenth of that of light (about 30,000 km. per second), the speed varying with the applied voltage which is causing the discharge. We shall see later (p. 607) that electron speeds higher than 290,000 km. per second have been observed in the beta rays; that is within 1% of that of light.

The ratio of charge to mass. The next steps are to discover the charge borne by each particle, and its mass. Thomson, using the apparatus already described (Fig. 28–4), measured the deflection produced by the magnetic field alone. This must be caused by the field always exerting a force perpendicular to the motion, and of constant amount; which requires that the path must be circular (p. 70). The radius R of the circular path in the field can be calculated from the observed deflection on the screen S, and the dimensions of the apparatus, by simple geometry. The centripetal force is mv^2/R; the magnetic force is Hev. Hence, along the circular part of the path

$$\frac{mv^2}{R} = Hev,$$

from which it follows that

$$\frac{e}{m} = \frac{v}{RH},$$

(E–M units again used), and, as the velocity has just been found, the right-hand side of this equation is now known.

The ratio of the charge to the mass given by the last equation is not all that we wish to know about these cathode-ray particles. We must find each quantity separately. This cannot be done from these experiments, and has not been done as yet from any experi-

ments on cathode rays alone. Before going further, however, we should extract from the numerical value of e/m (1.768×10^7, expressed in E–M units) all the information that we possibly can.

The phenomena of electrolysis yield the result that each hydrogen ion carries a certain small charge which appears likewise on other sorts of ions, whose value was given (p. 363) as 4.77×10^{-10} E–S units, or 1.59×10^{-20} E–M units. The mass can be shown to be almost exactly the same as that of a hydrogen atom, which is $1.008/6.06 \times 10^{23}$ or 1.66×10^{-24} grams. Hence the ratio of the charge E to the mass M for a hydrogen ion is

$$\frac{E}{M} = \frac{1.59 \times 10^{-20}}{1.66 \times 10^{-24}} = 9580 \text{ (E–M units).}$$

Thus the ratio of charge to mass for the cathode-ray particles is

$$\frac{17{,}680{,}000}{9580} \cdot \frac{E}{M},$$

or about 1840 times larger than for the hydrogen ions. This might be because the charge on the cathode-ray particle is larger than that on the hydrogen ion, or because its mass is smaller. The facts lead us to the second of these alternatives.

The charge borne by the electron. The evidence from electrolysis strongly indicates the existence of the electron as an indivisible electric charge, and shows that it occurs in a great variety of materials. The oil-drop experiment (p. 308) led to the same conclusion. Millikan used drops of oil (non-conducting) and of mercury (conducting) and measured the charges picked up by them from atmospheric ions, as well as the charges acquired by friction by the drops themselves in their formation at the nozzle of the atomizer which he used. All these experiments yield the same numerical value for the charge on the electron, and indicate its widespread occurrence. It is most natural, therefore, to conclude that the charge borne by the particles in cathode rays is identical with that found in other cases. This conclusion is strengthened by Thomson's discovery that when the metal of the cathode and the nature of the residue of gas left in the tube are changed identical cathode particles are still produced. By other methods (for instance the study of the Zeeman effect, p. 580) strong indications have been found that electrons even while inside an atom have almost exactly the same value of e/m, and therefore presumably the

same charge as they do when they are free. Thus the available evidence all points to the constancy and universal distribution of electrons.

The mass of the electron. The values of the ratio of charge to mass considered above force us now to an important conclusion. Since

$$\frac{\frac{e}{m}}{\frac{E}{M}} = 1840,$$

and $e = E$ it must follow that $m = M/1840$; or ***the mass of the electron is*** $1/1840$ ***of that of the smallest atom.*** This discovery, due to Thomson, was the foundation of the electron theory, many of the conclusions of which we have already examined. From this beginning arose the "nucleus atom," the conception of the arrangement of the electrons about the nucleus, and many theories concerning the chemical, electrical, and optical behavior of matter.

Variation of mass with velocity. We are familiar with inertia effects in electrical circuits which we classify under the head of self-induction. Since a moving charge is equivalent to a current, we ought not to be surprised to find an amount of inertia associated with electrons when they are moving which is different from what they have when they are at rest. The experiments described above offer an opportunity of detecting this effect, but others are still better.

Electrons are thrown out from atoms of radioactive substances with speeds far greater than those attained in cathode rays. The value of e/m for these "beta" rays (p. 607) has been measured at different speeds and found to vary considerably, especially as the speed approaches that of light. If it seems reasonable to assume that the charge does not vary, we are forced to the conclusion that the mass increases with the speed. The electromagnetic theory indicates that the mass of an electron should become infinite at the speed of light, and the observed variations are in agreement with this theory throughout the range of speeds for which observations can be made. The changes of mass are, of course, imperceptibly small at all ordinary speeds.

The same theory also shows that the more concentrated a charge

is the larger its inertia becomes. If we are willing to make the assumption that the relatively great mass of atomic nuclei is due to their electric inertia, it must follow that they are very small indeed, some 10,000 times smaller in diameter than the atom itself. The evidence as to the size of nuclei obtained from the scattering of alpha rays (p. 609) leads to the same conclusion.

Whether all mass is due to electrical energy or not is a question on which the theory of relativity has something to contribute (p. 618).

Discharges at moderately low pressures. In the "vacuum" tube experiment already described (Fig. 28–2), the pressure was progressively decreased from its atmospheric value until it had reached a few millimeters only, and then the tube was filled with a glowing discharge. Wherever there is such a glow, the light is due to the recombination of gas ions, and when differences in color occur, they indicate the presence of ions of different sorts. Ionization is much more likely to occur in a low-pressure tube than ordinarily because the ions move a longer distance between collisions and the driving electromotive force has a longer time in which to accelerate them up to the speed at which collisions produce ions. This multiplication of ions by collision occurs so readily in low-pressure gases that they may contain as many as 10^{12} ions per cubic centimeter (including unattached electrons). Such gases are very properly listed as conductors of electricity.

The mechanism of the discharge though very complicated is now fairly well understood. Electrons shoot off from the surface of the cathode in the form of cathode rays because positive ions are attracted to it and strike it vigorously. In a gas at moderate pressure they go only a very short distance, the thickness of the cathode dark space. In this little journey the electrons are being accelerated, till they acquire speed enough to ionize the gas particles they hit. The rate of fall of potential along the tube is greatest in the space immediately next to the cathode. This is usually referred to as the "cathode fall." In other parts of the tube the changes of potential are slower. The details are complex.

Positive rays. The existence of the positive ions just referred to can be shown in a tube of the form shown in Fig. 28–5. Here C is the cathode and A the anode. C is perforated so that if there are ***positive*** particles which are attracted toward it from the direction of A, some of these may shoot through the holes and continue be-

yond into a space comparatively free from any electric field, in which they travel as positive rays. As they came through little channels provided for them in the anode, they were at first called "canal rays." Such streams of positive ions are very faintly luminous and can be shown by deflection experiments, similar to those on cathode rays, to consist of atoms or molecules bearing single or multiple charges.

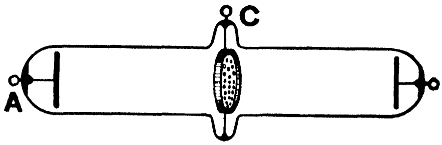


Fig. 28–5

J. J. Thomson in 1907 made a discharge tube of the form shown in Fig. 28–6 in which the positive rays were limited to a narrow beam by being made to pass down a narrow straight hole in the tube T. They then passed through a space in which there was a strong electric field deflecting the rays horizontally, and simultaneously a magnetic field deflecting them vertically. The results of these deflections showed themselves on a photographic plate against which the positive rays hurled themselves at the end of their journey. From the magnitude of the deflections the velocity

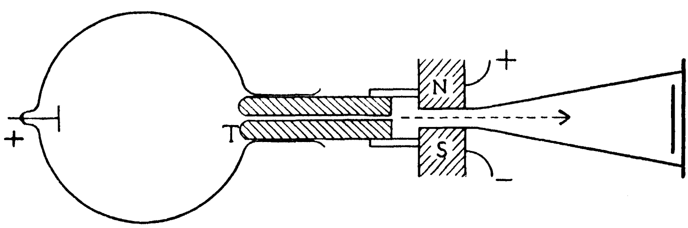


Fig. 28–6

and the ratio E/M for the ions could be obtained and thus the nature of the ions could be identified.

The mass spectrograph. Isotopes. The results which were obtained with the apparatus just described proved to be so important that the work was continued by Aston [1] with an improved form of apparatus known as the ***mass spectrograph,*** (Fig. 28–7), capable of yielding results of higher precision. In Aston's form of the instrument the positive rays, restricted to a narrow beam as before, are first bent by the electric field and then given a still greater curvature by the magnetic field now acting in the opposite

[1] Dr. F. W. Aston of Trinity College, Cambridge; author of "Isotopes," 2nd edition, 1924 (Edward Arnold & Co.) in which a full account of this work and its connections is given.

direction. By a suitable arrangement of the apparatus and choice of field strengths, all the rays due to one sort of ion are brought to a focus upon a small spot on the photographic plate, and the same spot is reached no matter at what speed the particles are moving. As there always is a considerable range in speed among the particles in positive rays, there was a spreading of the image in Thomson's apparatus, which is now avoided, with the result that fainter impressions can be recorded than before.

Thomson found photographic images produced by the positive ions of H, H_2, N_2, O, O_2, CO, and many others. The atoms of several gases were found to have the masses yielded by chemical methods, but when the inert gas ***neon*** was tried, there was no spot on his plate corresponding to its atomic weight 20.2, but instead there appeared a strong record of atoms of atomic weight

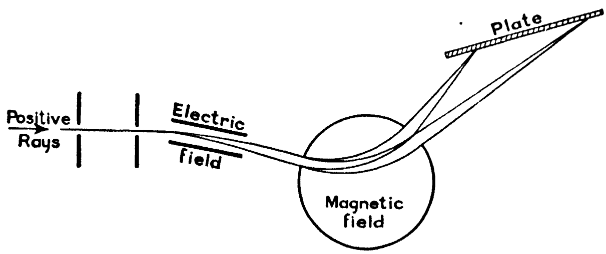


FIG. 28–7
A sketch illustrating the action of the mass spectrograph

20 and a weaker one corresponding to a weight of 22. Thus it appeared that there are two sorts of neon, mixed in such a proportion as to yield the usual atomic weight as an average. These are identical in chemical nature, and therefore, as we shall see, in the arrangement of their external electrons, but they differ in the atomic nucleus (p. 609) where the mass is concentrated. Elements with the same number of electrons but different masses are called ***isotopes;*** the word implies that they occupy the same place in the table of the chemical elements. Isotopes have identical chemical properties and are almost impossible to separate from one another.

Aston found many more isotopes. Chlorine, for instance, consists of two, with atomic weights 35 and 37, yielding the average value 35.46 which is its accepted "atomic weight." It is worth noting that the atomic weight of chlorine is thus ***not*** the weight of any chlorine atom, but only an average of different numbers.

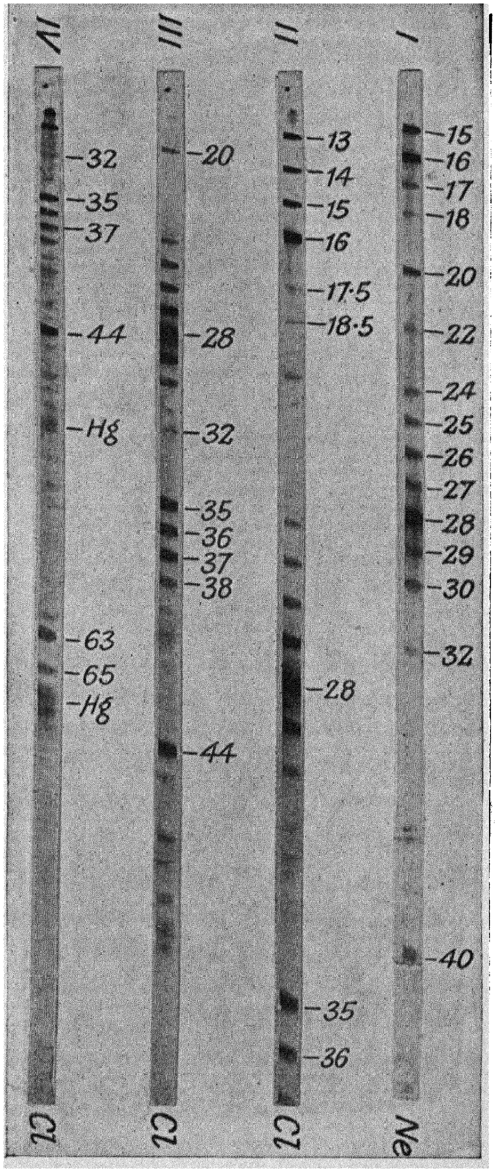


FIG. 28–8

From Aston's "Isotopes" (Edward Arnold and Co.).

The Periodic Table (Appendix) gives the chemical elements, with their isotopes, as determined by many observers. The phenomena of radioactivity led by independent methods to the discovery of many other isotopes, as explained below (p. 611).

Figure 28–8 shows one of Aston's photographs, taken with the mass spectrograph, in which are seen spectra taken with (I), a mixture of neon and carbon monoxide (CO), and (II, III, and IV) phosgene ($COCl_2$). In I the two neon lines, 20 and 22, are visible, the CO line at 28 is strong, and the others are due mainly to compounds of C, H, and O. In the other spectra one sees the Cl lines at 35 and 37, HCl at 36 and 38, CO_2 at 44, doubly-charged Cl at 17.5 and 18.5, trebly charged Hg at 66.6, and many lines also found in I.

In the case of gases it is not difficult to obtain positive rays of their atoms in discharge tubes in the manner described. With solids it has been possible to generate them by means of electrically heated filaments coated with salts of the element desired. Positive ions of the element are thrown off and may then be treated in the same manner as gas ions.

PROBLEMS

1. An electron has a charge of 4.77×10^{-10} E.S. units and a mass of 9.0×10^{-28} grams. It is emitted by a hot wire which is acting as cathode in a vacuum tube, and is accelerated through a potential difference of 20 volts (300 volts = 1 E–S unit). Find the final energy and velocity of the electron, assuming its speed to be due entirely to the accelerating field. If it then has just speed enough to ionize a certain atom, find how many ergs of work are necessary to do this.

2. A discharge tube, such as that in Fig. 28–1, is highly exhausted and then has most of its fall of potential localized at the surface of the cathode. If an induction coil furnishes 100,000 volts to the tube, what is the work done on an electron in its passage through this potential difference, and what will be the velocity acquired by it? Would m actually remain constant?

CHAPTER 29

ELECTRON TUBES AND THEIR APPLICATIONS

Discharges with hot electrodes. Thermions. Electrons seem to be more or less free to move inside a conducting wire, but to have very great difficulty in escaping from it. In recent years it has been shown that in a high vacuum they can be pulled out by an extremely strong electric field, but ordinary high fields cannot do it. They can be freed by the impact of rapidly moving atoms, as happens in the generation of cathode rays, for instance, and they are thrown off from the surfaces of some metals when ultra-violet light falls upon them (p. 596). But the easiest way for electrons to escape comes when the wire is heated. Apparently the thermal agitation in a red-hot wire so upsets the usual restraining influences that electrons can "evaporate" freely, and in a tube containing practically no residual gas (the pressure being as low as 0.000001 mm.) a hot wire becomes surrounded with a thin atmospheric film of electrons, quite invisible, but ready to do wonderful things if properly urged. These escape also even when the pressure is not so low, but then they usually associate themselves before long with atoms, which are then appropriately called ***thermions,*** and the resulting phenomena are more complicated than they are in a vacuum.

Edison (1883) noticed that a hot wire in a vacuum tube furnished with a third electrode, as in Fig. 29–1, (two being used for the ends of the filament) would give a current through the vacuum if the hot wire were connected to the negative terminal of a battery, and the third electrode to the positive; but no current would flow

if the direction were reversed. We now know that this was because the electrons which are needed to carry this current cannot escape from the cold electrode. The current in a vacuum tube does not follow Ohm's law in the ordinary way, or, in other words, the resistance cannot be regarded, even approximately, as a constant.[1] The electrons can escape at a certain rate, fixed by the temperature. Each electron carries with it a definite charge, and there are practically no gas ions present to carry any current because the pressure is too low. Hence if the E.M.F. on the tube is raised, the current rises according to Fig. 29–2, eventually reaching a "saturation" value when all the electrons that evaporate from the hot wire travel over and deposit their charges as part of the current. Unless more electrons can be set free, no more current can flow no matter how large an E.M.F. is applied.

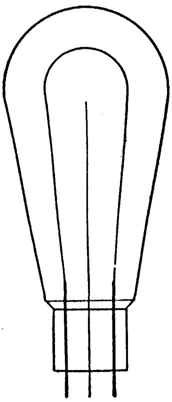


FIG. 29–1

Under ordinary circumstances the space about a hot wire becomes full of electrons. These escape at a fixed rate for a given temperature, and are many of them recaptured by the wire itself. In the presence of an electric field, as in Edison's experiment, many move in response to it toward the positive electrode. Thus at any instant the space near the wire may be regarded as having a volume charge, called ***space charge.*** This, being negative, tends to reduce the attraction of the positive terminal for the electrons as well as to drive escaping electrons back. On both accounts the space charge seriously limits the number that reach the positive electrode and thus

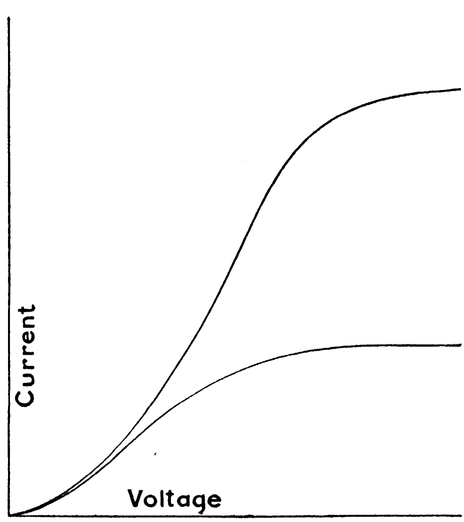


FIG. 29–2
Electron current at two different temperatures

[1] There are electron tubes so arranged that within certain limits the "resistance" is negative, i.e., the greater the E.M.F. the less the current.

reduces the amount of current that can be obtained through any such tube.

Electron tubes or valves. In Fig. 29–3 the dotted circle is supposed to represent a glass tube whose walls and contents have been highly heated and exhausted, so that a very high vacuum (of the order of 10^{-5} or 10^{-6} mm.) exists in it. It is usually coated on the inside with a thin film of a chemically active metal (e.g., magnesium) which will combine with any gas that may be given off in the tube while it is being used, so that the vacuum may thus be maintained unimpaired. Inside this tube is a filament of tungsten, F, heated by an electric current from a low-voltage battery A. Usually the tungsten has some thorium in it which diffuses out to the surface when the wire is very hot and forms a layer known to be just one atom thick all over the outside of the wire. This "thoriated" tungsten wire emits electrons much more freely than pure tungsten. These electrons form a swarm in the space about the hot wire, as already stated.

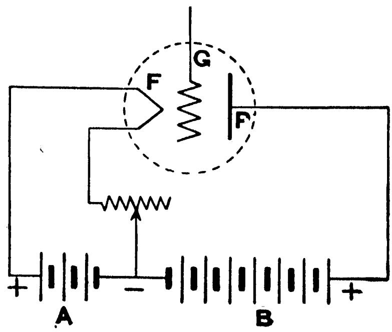


FIG. 29–3

A three-electrode tube in a suitable circuit

The tube contains a plate P charged positively with respect to the filament by means of a battery B, which consists of many cells in series. This creates an electric field serving to draw electrons over to the plate, thus producing what is called the plate current, I_p. This can flow in only one direction. Obviously, if the "B battery" were so connected that the plate was negative, it would be trying to draw electrons out of the plate, which is impossible because it is cold. Hence this tube is a ***rectifier,*** acting like (and often called) a valve, which allows a flow in one direction only.

If this were all, the flow of current would be seriously limited by space charge. The third electrode is connected to a grid G, which is a mesh of fine wire situated between the filament and the plate. (Actually both the grid and the plate usually surround the filament.) The grid is usually kept at a negative potential with respect to the filament so that it will not attract electrons to itself. Small changes in its potential then exert an extraordinary effect on the current to the plate. When the potential of the grid is only

slightly negative, the positive potential of the B battery overcomes its influence on the electrons and draws a current over to the plate. But a slight increase in the negative charge on the grid tends to reduce this current and may be enough to stop it almost entirely. Thus a large current to the plate may be controlled by small changes in potential on the part of the grid. One is reminded of the tip of a dog's tail, which may be wagged through a large amplitude by a relatively small motion on the part of the muscle doing it.

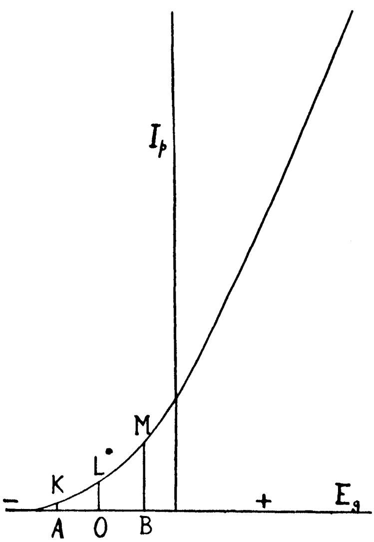


Fig. 29–4
Characteristic curve of a three-electrode tube (a "triode")

Use of the tube as a detector. Such a tube may be studied by making a series of measurements with a given plate voltage showing how the current I_p to the plate varies with the difference of potential E_g between the grid and the filament. The "characteristic curve" in Fig. 29–4 shows the results.

Let us suppose that the potential of the grid is made negative as at point O, and is then made to oscillate through a small range, such as AB. This happens, for instance, if it is connected to a radio antenna and electric waves are coming in from a distant source. The plate current increases on the positive half of an oscillation from a value OL to the much greater value BM. In the negative half of the oscillation it diminishes to AK. Hence the

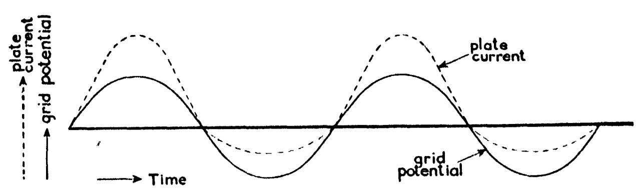


Fig. 29–5
Oscillations in a tube used for detection

oscillation produces an unsymmetrical change in the plate current, as shown in Fig. 29–5. The average effect of the changes in grid potential is thus to increase the current to the plate. If a group,

say 50, of such oscillations occurs 500 times a second, each group being followed by a period of silence, as in Fig. 29–6, there results a succession of forward impulses of current which produce in a telephone a musical tone of a frequency of 500 per second. The telephone because of its inertia cannot follow the individual oscillations in a group, which are perhaps at the rate of 1,000,000 per second. The sending station may create such groups of oscillations by a regular succession of sparks (p. 436) at the rate of 500

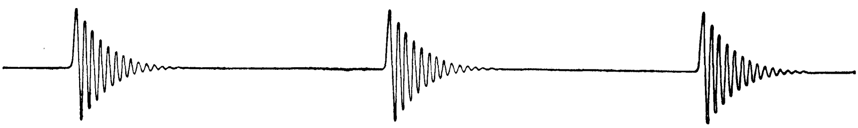


FIG. 29–6
The oscillations in a series of spark discharges

a second. Fifty of these would make a short signal in the telephone; 200 or so a long signal, and the listener might thus receive the dots and dashes of the telegraphic code signals.

This is one of the processes by which ordinary wireless telegraphy can be received. The reception of the more modern continuous-wave signals is referred to below (p. 460).

In one way the action of a detector tube is not quite satisfactory. When the oscillations in the grid potential, produced by the incoming waves, are doubled in amplitude, the resulting plate current is not exactly twice what it was before; or, in other words, the response of the tube is not in exact proportion to the incoming oscillations. This introduces a certain amount of distortion (p. 465) in the reception of speech and music in radio telephony, which has the effect of slightly changing the quality of the sounds reproduced, but this distortion may be made almost negligible.

The electron tube as an amplifier. The average potential of the grid may be maintained at a certain negative value ("biased") by means of a third battery ("C battery"), as indicated by the point C (Fig. 29–7). The plate current is now given by CG, and the point G is in a portion of the "characteristic curve" of the tube which is practically a straight line. If the grid, thus biased, is made to change through a small range of potential AB by being connected to some outside source of oscillations, the plate current will oscillate in proportion, as in Fig. 29–8; its changes are enlarged but faithful copies of the changes in grid potential.

Under these conditions the tube acts as an ***amplifier.*** Incoming signals received on the grid are amplified in the plate circuit. The usual custom is to employ several tubes for this purpose in series, in which case the amplified oscillations in the plate circuit in the first tube are led to the grid of the second, and produce much greater oscillations in the latter's plate circuit; and so on, if desired, through several ***stages of amplification.***

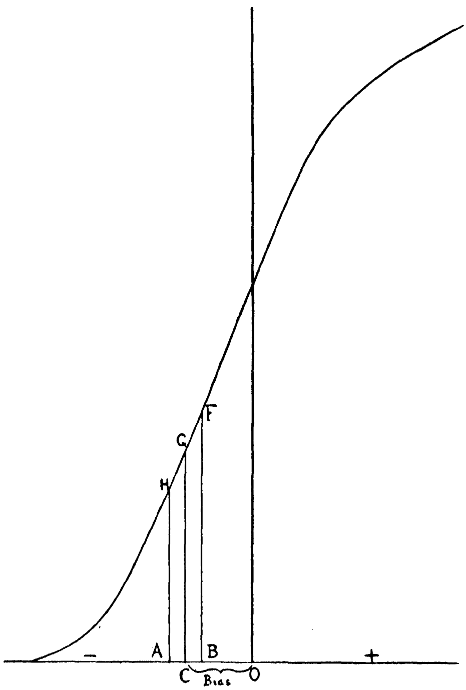


Fig. 29–7

The negative bias on the grid is essential. If the grid were positive, a strong current of electrons would rush over to it instead of to the plate. If a series of tubes is used for successive amplification, the last one, usually called the "power" tube, may have its grid oscillating through a potential range of several volts. To keep the quality of the reproduction satisfactory, these oscillations must all lie on the straight portion of the characteristic curve, and the grid must remain negative throughout. Hence power tubes must have a large negative bias, sometimes as high as forty volts.

This use of a three-electrode tube is of inestimable advantage in many sorts of physical research. It has been applied to the magnification and measurement of the electrical effects in the eye accompanying vision; to the measurement of extremely minute quantities of heat radiation, artificially interrupted so as to create oscillations; to the detection of individual atoms (p. 612), as well as to many other scientific and practical uses. Perhaps the latest of these has been to make audible the activities of a fruit fly larva inside an apparently good grapefruit. Of course, its best-known application is in radio telegraphy and telephony. It seems very nearly magical to be able to receive the excessively feeble electric

apparatus, though it is beyond the scope of this book to go into elaborate details of this sort. In Fig. 29–9, R is a receiving antenna or "aërial" connected through a coil Q to the earth at E. L is a coil which is so close to Q that oscillations in Q induce others in L. These, in turn, charge the condenser C, which can be varied in capacity (as signified by the arrow across it in the figure). Oscillations may start in the system LC if the inductance and capacity have proper values (p. 433). By tuning C this circuit can be set in resonance with the oscillations in RQE, which have been picked up from a distant source. These oscillations will charge the grid G of the tube. The oscillations of the charge on G will cause the electrons from the hot filament F to flow in greater or less numbers than usual over to the plate P. The filament circuit includes the A battery, the filament, and an adjustable resistance which regulates the current through the filament, and hence its temperature. The plate circuit includes the B battery, the filament, the electrons in the tube, the plate, and the telephones (or the telephone condenser TC). The electrons flow from F to P in gusts, which follow one another so rapidly that the inductance of the telephone would block their oscillations. Hence the condenser TC is put in to enable these to cross back to the B battery by that route. In so doing they charge this condenser, which then in a more leisurely way sends one surge of current through the telephones for each group of oscillations, and it is the repetition of these surges that produces a tone audible there. (The usual twisted cord of the telephones has enough capacity to act in place of the telephone condenser.)

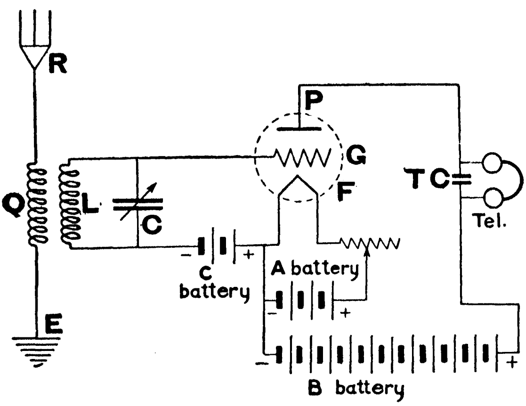


FIG. 29–9
A radio receiving circuit

Thus the incoming waves merely have to charge the grid, which then acts, like the handle of a faucet, to turn on or off a large stream in the plate circuit, the power for which comes from the B battery.

In the most recent forms of receiving sets for use in districts

supplied with alternating current there are no batteries; everything needed is supplied from the lighting circuit. There are, however, devices which take the place of batteries (p. 462), so that the description of the action of the set can be given in the same words as above.

Regeneration. Oscillating Circuits. The circuit in the last figure may be modified by the insertion of a small "tickler coil," T, (Fig. 29–10) in the plate circuit, which is placed very near to the coil L. When the large pulses of current in the plate circuit (it is hardly fair to call them oscillations, as the current is always in the same direction) pass through the coil T, these may "feed back" some of their energy by induction into the oscillating circuit LC, so as to increase the oscillations there. Naturally, the winding in T has to be in the right direction to get this favorable result. When LC begins to oscillate more vigorously, this acts on the grid and creates still larger effects in the plate circuit, and so on. Thus, if the inductive influence of T on L can be varied (for instance by turning T), a correct setting may yield an enormous increase in the sensitivity of the detector, and hence will greatly increase the loudness of the sounds in the telephone without altering their nature. The incoming vibrations still settle what shall happen in LC, and hence throughout the apparatus, but the B battery energy "regenerates" these oscillations and magnifies them perhaps a hundredfold, or more.

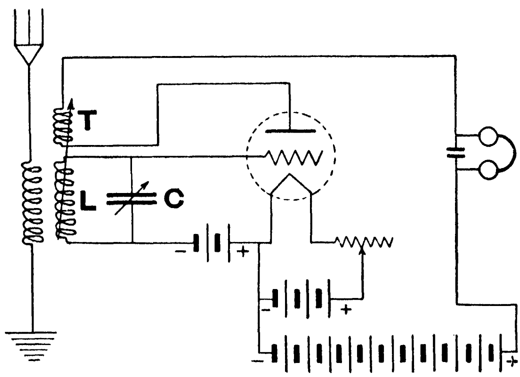


FIG. 29–10
A regenerative circuit

If the effect of T on L is made still stronger, so much energy is fed from the B batteries back into L that the incoming signals are not needed to start the oscillations, and the system will begin oscillating at the rate which is natural to the oscillating circuit LC. Such ***oscillating circuits*** (without the antenna) are extremely useful, and are really self-starting, since small variations in the battery current are enough to upset their electrical equilibrium, and the

"feed-back" then generates the oscillations. Alterations in the variable condenser, or in the inductance of L, will then alter the time of oscillation according to the relation $T = 2\pi\sqrt{LC}$ (p. 417). If the inductance L and the capacity C are both large, the rate of oscillation may be made slow enough to fall within the range of audible sound frequencies, and such an arrangement is known as an "audio-oscillator." It is often used with a loud speaker to produce pure tones over a wide range of frequency and volume for acoustical experiments.

Oscillations in the range of 15,000 to 200,000 per second are easily created and may be used to produce ***supersonic*** vibrations (p. 266). These oscillations are led into the plates of a condenser between which is a plate of quartz (as in Fig. 20–15, p. 319). The quartz has certain natural frequencies like any other elastic solid, depending on its elastic properties and its dimensions. When the electrical oscillations are tuned to these natural frequencies, vigorous vibrations are set up by piezo-electric action, which then pass out into the air as supersonic waves. They may be made still stronger by using several quartz plates of the same thickness to shake a large steel plate to which they are attached, and thus emit a considerable amount of energy. Such devices are used under water to create beams of short waves which can be directed like searchlights.

Oscillating circuits may also be used at the so-called "radio frequencies" (roughly 20,000 to 100,000,000, overlapping the supersonic range). Amplified incoming waves may be made to produce "beat-tones" by combining with oscillations created in the receiving apparatus. Thus, if waves at 150,000 frequency are being received and oscillations at 150,500 are being generated in the receiving set, these produce 500 beats per second (p. 259), which can under proper conditions be heard as a note of frequency 500 in a telephone receiver. This is the "heterodyne" method of reception which is much used in continuous-wave wireless telegraphy.

Radio sending circuits. The most important use of oscillating circuits is to send out continuous waves. This method is now commonly used for wireless telegraphy, and universally used for wireless telephony or radio broadcasting. The diagram already shown (Fig. 29–10, p. 459) for a receiving circuit needs little modification to turn it into a sending circuit. Oscillations are

generated as before in the circuit LC, and these create by induction corresponding oscillations in the antenna circuit, to which they are tuned to resonance. The antenna circuit sends off the energy of these oscillations in the form of electromagnetic waves whose length λ depends on their velocity V and their frequency f, according to the familiar formula $V = f\lambda$ (p. 242). As the velocity of these waves is 3×10^8 meters, the common range of wavelengths used at present in broadcasting (200 to 600 meters) corresponds to $\frac{300{,}000{,}000}{200 \text{ to } 600}$ or 1,500,000 to 500,000 vibrations ("cycles"), or 1500 to 500 "kilocycles" (kc) per second.

When a receiving set is oscillating, it may send out waves which, sometimes unintentionally, create disturbances in other sets a few miles away. To send signals hundreds of miles, however, requires much more power than the one or two watts usually furnished by the B battery. Power tubes are now made up to 100 kw. in size, water-cooled to carry off the heat developed within them; they have as much as 20,000 volts in the plate circuit.

Radio telephony. Carrier waves. The problem of transmitting speech by means of electromagnetic waves has been solved in a way quite different from what one might expect. One would naturally suppose that the air vibrations could be transformed into electric vibrations of the same form as they are in a telephone, that these might be sent out as waves, and be caught by the receiver and made to move a telephone diaphragm. While such a process might be used over a distance of a few feet, it would be difficult to put much power into it for long-distance transmission, and since one set of waves could not be "tuned out" from another, confusion would result if two stations were broadcasting at once.

The successful method now in use involves a quite different idea, that of a ***carrier wave,*** which is "modulated" by the voice. Electric oscillations are produced in a circuit at a rate of, say, 1,000,000 per second. These are as indicated in Fig. 29–11*a*. The voice is then made to alter the amplitude of these, in proportion to the amplitude of its own vibrations (which are shown in Fig. 29–11*b*), the result being indicated in Fig. 29–11*c*. If the voice makes one complete vibration in 1/200 sec., there must then be 5000 of the rapid vibrations during the time of one sound oscillation. (The individual vibrations would not show if they were drawn on this scale in the proper numbers to fit this example.)

The rapidly oscillating wave is called the ***"carrier wave."*** Its frequency may have any (high) value we please; its amplitude follows the vibrations of the sound which it "carries." Figure 29–11*d* shows the modulated radio wave as received in the "detector" tube, almost all the oscillation in one direction being checked, and the waves reduced to rapid pulses in the other direction occurring in groups of varying vigor. (A modern power detector rectifies more completely than Fig. 29–11*d* indicates.) The receiving set treats these in the manner explained on page 454, that is, the telephone cannot respond to the individual pulses, but follows their general trend, and thus reproduces the form of the original sound wave, as in Fig. 29–11*e*.

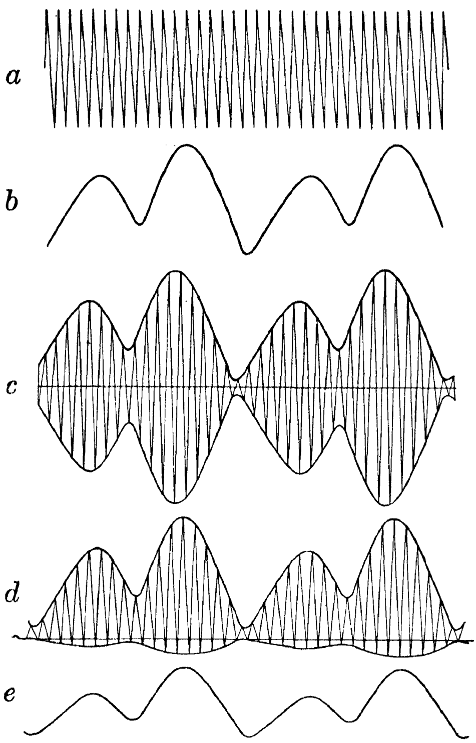


FIG. 29–11

The process of modulation involves first the creation of electric oscillations similar to those of the voice, which is done in much the same way as in the telephone transmitter. These oscillations are amplified and then imposed upon the rapidly oscillating circuit through induction. They thus combine with them in such a way as to produce the desired modulated waves, such as those in Fig. 29–11*c*.

Radio sets without batteries. The most convenient sort of radio receiving set is one that derives all its electrical supply from the usual house lighting (A.C.) circuit. The batteries, *A*, *B*, and *C*, are "eliminated," though, as their functions must still somehow be performed, it is a case rather of substitution than of elimination.

If the filament of a tube is heated with alternating current, the temperature fluctuates somewhat with the current, and the receiving set gives a humming noise due to the 60-cycle period of the alternations. The simplest way of

overcoming this difficulty is to enclose the hot filament in a light close-fitting conducting cylinder, which after a few seconds becomes red-hot from the heat of the filament within it, so that its surface then emits the necessary electrons. The heat-capacity of the cylinder is so great that it does not fluctuate in temperature with the alternations of the current. This then eliminates the humming noise and furnishes a steady source of electrons without requiring any A battery.

The process of deriving high-voltage direct current from alternating, to eliminate and act in place of the B battery, is more complicated. A transformer is used to raise the voltage (if necessary) and then the current passes through a "rectifier" (see below) which transmits only one-half of each cycle of current, thus changing it into pulsations of direct current. Finally a group of condensers and inductances, whose action cannot be considered here in detail, smooths out the irregularities of the current into a steady flow. Hence a supply is furnished similar to that which would be given by a B battery.

The bias otherwise furnished by the C battery can be derived from part of the supply coming from the "B-eliminator."

Rectifiers. Rectifiers are devices which allow current to pass freely in one direction, but with difficulty or not at all in the other. The tubes already described (p. 452) in which the current is carried by electrons from a hot terminal to a cold one are perfect rectifiers. In large sizes these are used under the name of "kenotrons." Another such device which is new and interesting consists of a film of copper oxide on metallic copper. For some reason, electrons can flow readily from the metal into, and through, the oxide film, but very little in the opposite direction. A cell containing a plate of aluminum and a plate of another metal (e.g. lead) dipping into a solution (such as borax) shows a similar rectifying action at the surface of the aluminum; but any wet cell is less convenient than the copper-oxide type of rectifier.

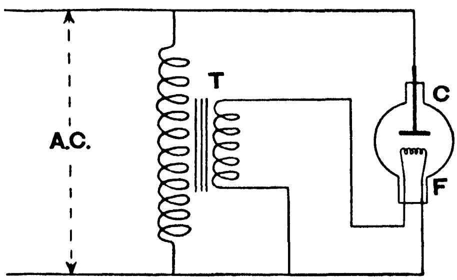


FIG. 29–12

The most common use of rectifiers is to charge batteries from an A.C. supply, or to furnish direct current for electrolytic cells. This is very frequently done by using a *"tungar"* (tungsten-argon) *rectifier.* In this device (Fig. 29–12) the tube R contains a heated tungsten filament F, and a carbon electrode C, and is filled with an inert gas (argon) at reduced pressure. One side of the filament is connected to one terminal of the A.C. supply, and the electrode C to the other. There is also a step-down transformer T, which supplies low voltage alternating current for heating the filament. During the time that C is positive and F negative, electrons can escape from F and flow over to C, but during the other half of each cycle the electrons cannot escape from F on account of electrostatic repulsion, and none can get out of C because it is not heated. Hence the current goes in gusts, during half of each oscillation, with no current in the intervals between. The inert gas

is ionized by the electrons and thereby greatly increases the amount of the current produced. The fact that the current is not uniform makes no difference in the uses to which this device is put.

"Wired wireless." The methods of radio telephony may advantageously be applied to telegraphy and telephony with wires. For instance, a carrier wave, say of 50,000 frequency (much too high to be audible if it were a sound wave), can be sent over a telephone line, and its amplitude can be "modulated" by a telephone transmitter so that it "carries" the voice vibrations with it. At the far end of its journey it meets a set of *electrical filters*, a remarkable modern invention composed of groups of condensers and coils with inductance. Filters can be made of various sorts; their peculiarity is that they let vibrations of certain frequencies through, but stop others. If one particular filter lets these 50,000-cycle vibrations through to a certain line, while other filters prevent them from going anywhere else, there may be at the far end of this line a receiving set (practically a radio-telephone set) where the voice may be heard. While this conversation is being carried on by means of the 50,000-carrier wave, a second conversation can be sent over the same line, sorted out by an appropriate filter into a different end-connection, and received inde-

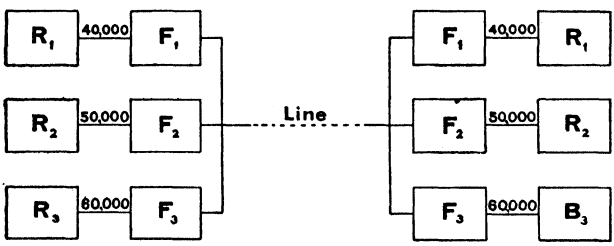


FIG. 29–13

pendently, as indicated schematically in Fig. 29–13. Thus a single telephone line may be transmitting an ordinary telephone conversation, half a dozen others each on its own carrier wave, as well as some ten sets of telegraph signals, all on carrier waves of different frequencies, no one of these interfering with any other. Each frequency range is sometimes referred to as a "channel"; thus 45,000 to 55,000 would be a channel for the 50,000-carrier wave. If the channels are 10,000 cycles apart there is no overlapping, and they may begin at 20,000 and go up to perhaps 200,000, the process becoming inefficient at higher frequencies for several reasons. In actual practice the carrier wave is complex, and only a portion of it is sent over the wire, but such complications are beyond the scope of this book.

Loud speakers. It is quite a problem to devise a piece of apparatus which will faithfully follow the intricate variations of the currents in a radio set and reproduce without alteration the sounds to which they correspond. The mathematical theory of this problem has recently guided experimenters to an almost complete success.

Among many others, two common schemes might be mentioned. In one (the so-called "dynamic" speakers) a very thin coil (Fig. 29–14) attached to a

light cone is placed in the concentrated magnetic field of a powerful but small electromagnet. Feeble alternating currents in this coil make the coil vibrate, and thus the cone also, which in turn passes the vibrations on to the air in considerable volume. The amplified oscillations from the radio set are of course sent into the little coil. Another common type (Fig. 29–15) receives the oscillating currents in a fixed coil whose varying magnetic field attracts or repels a light, pivoted permanent magnet, the motions of which shake a light but stiff paper cone of flat shape and large size.

The chief difficulties with loud speakers arise from the fact that they are themselves elastic mechanical systems with natural periods of vibration of their own. If the incoming vibrations coincide in frequency with one of their natural periods, a very loud response results, but if they do not agree, the response may be very feeble. Hence if a mixed sound occurs, one of whose

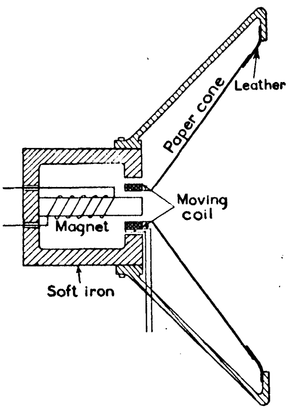


FIG. 29–14
A "dynamic" speaker

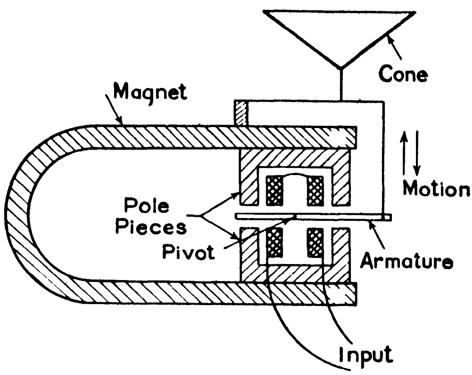


FIG. 29–15
A magnetic speaker

component vibrations is thus unduly amplified, the sound is said to be *distorted*, and the effect is unnatural.

Other troubles arise from the difficulty of making large bodies respond to rapid vibrations on account of their inertia. Hence the highest frequencies in the source may not be reproduced. For other reasons the lowest vibrations are commonly unduly weak. These defects are now practically eliminated from the waves sent out by broadcasting stations; that is, these waves "carry" an undistorted copy of the original sound at least over a range of frequency from 50 to 5000; but receiving sets and loud speakers only too often fail to do their part perfectly. Improvements in this respect have recently occurred and may be expected to continue in the near future.

Radio as applied to navigation. ***Radio beacons*** are a new invention which can tell an aviator whether he is on a certain course or not. A radio sending station at the airport toward which he is traveling creates vibrations in two large loops of wire at right

angles to each other and at 45° to the desired course (Fig. 29–16). One form of receiving device on the airplane has two reeds set into vibration, each by one of the waves from the two loops. These reeds vibrate equally anywhere along the straight course; but if the aviator deviates to one side, the radiating loop towards whose plane he deviates sends him stronger vibrations, which are indicated by a more vigorous vibration of the reed on that side. It is a property of a plane loop of wire acting as a radiating antenna that it sends strong waves out in its own plane and practically none in a direction at right angles. Hence anywhere within the 90° angle between the two loops the aviator can determine how to go in order to reach and keep on the straight course. Radio beacons can be operated from the point of departure or the place of arrival, and have already proved their worth in making straight steering possible in long ocean flights. At a landing field equipped with a radio beacon the loops might be turned around toward any approaching aviator who desired to have a course laid out for him. Actually, however, the loops are much too large to be capable of being handled in this way, but the same thing can be accomplished in other ways. Radio beacons are, of course, equally applicable to ordinary navigation.

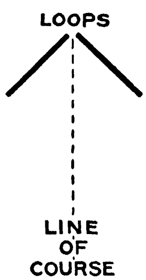


Fig. 29–16

A ***direction finder*** is a receiving loop which has the same property as the loop just mentioned; that is, it gives the strongest reception when its plane is aimed at the distant sending station. A direction finder on a ship can be set on two coast stations in succession, and thus the position of the ship can readily be found. Or, if the ship has no direction finder, but the shore stations have them, the ship may ask the shore stations to report its direction from each one, and hence find its own position.

A ***radio altimeter*** is a device which determines the height of an airplane above the ground. In one form of this apparatus the plane sends out waves of constant length which are reflected from the surface of the earth. Thus standing waves (p. 255) may be formed, with the surface as a node, and when this happens, more energy is required, and the reaction on the sending set is observable. The magnitude of this reaction changes with the number of wave-lengths that are formed in the set of standing waves, and thus gives some information about the height of the plane. This device,

or some other more or less like it, should make safe landings possible in fog.

"Static" and other disturbances. Fading. Listeners to radio stations some distance away are often troubled by noises which cannot be "tuned out," and which must therefore be of a non-periodic nature. These are called "static" and are explained as due to electric discharges in the atmosphere, perhaps lightning strokes, hundreds of miles away, or quiet but varying glow discharges nearer to the observer. These cannot be eliminated in ordinary apparatus, though an ingenious invention involving a receiving apparatus several miles long, called a wave-antenna, has made it possible to aim the receiving apparatus at the sender, so to speak, (very faint signals being received from any but the right direction) and at the same time to reduce the annoyance from static effects. This device is especially useful in very long-distance reception.

Other common disturbances arise from the turning on or off of electric lamps near the receiving apparatus, which produce inductive clicks in the telephone; or from the sparks in near-by automobile engines, or at the brushes of electrical generators, elevator motors, electric refrigerator motors, etc.

Fading is a phenomenon noticed with moderately distant stations, especially with those one to two hundred miles away, whereby the signals become faint at times and then recover their strength, the change in some cases occurring every minute or two and in others very rapidly. This effect has been shown to be due usually to the combination of ***two*** waves from the distant source, one of which has come rather directly along the surface of the earth, while the other has traveled upward to the conducting (ionized) layer in the high levels of the atmosphere (p. 446), and there has been bent and returned to the earth again. The erratic features of the behavior are due to variations in the condition of the upper air, so that these two waves sometimes agree in producing a large effect and at other times mutually destroy each other. The ground wave appears, however, to die out at great distances, and the fading that is still observed in these cases must be explained otherwise. The direction of the electric vibrations in the incoming waves is known to vary, and this produces changes in the amount of energy picked up by the receiving antenna which may furnish a clue to the explanation.

LIGHT

CHAPTER 30

QUANTITY, NATURE, AND SPEED OF LIGHT

Intensity and illumination, 468; variation of illumination with distance, 469; the photometer, 470; color effects in the photometer, 471; the flicker photometer, 471; practical illumination requirements, 472; recent improvements in light sources, 473; theories of light, 474; simple diffraction effects, 474; diffraction around a round obstacle, 476; Huygens' principle, 477; early measurements of the speed of light, 477; Michelson's experiment, 478; speed of light in other media than air, 479.

PHOTOMETRY

The human eye is a marvelous piece of mechanism, but in one important respect it is defective. It supplies us with no way of measuring the strength of its sensations. We cannot say that one lamp produces a sensation in the eye which is twice as strong as that given by another. We even have some little difficulty in making sure when two light sensations are of equal strength, especially if there is a slight difference in color in the two sources. Nevertheless, we ought to know how much light is given by artificial sources, how much we need for different purposes, and how this can most economically be supplied. Ways of getting such information as accurately as possible must be considered.

Intensity and illumination. The amount of light falling on a surface depends on the intensity of the source and its distance, and, if the surface is not facing the source, upon the angle also. The ***intensity*** of the source is measured in terms of candle power, as candles were the universal source of artificial light at the time this subject was first studied. The amount of light given out by a candle depends on the material of which it is made, the size of the wick, and the height of the flame. In the early days candles were made from a whale-oil product, called spermaceti, and the ***standard candle*** used to be defined in terms of this material. The interna-

tional standard candle is now defined (in England, France, and the United States) in terms of certain standard electric lamps, the exact specifications of which need not concern us. They are guaranteed to give a certain candle power in a definite direction when run with a certain current or at a certain voltage. Secondary or working standard lamps are prepared from these.

The unit of ***illumination*** is that given by one standard candle at a foot's distance (a "foot-candle") or at a meter's distance (a "meter-candle").

The latter unit has also received the name of the *lux*. In addition, workers in this field use a number of other terms with which we shall not be especially concerned, but they are given here for reference. The *lumen* is the amount of light (the "luminous flux") flowing through a square meter of area at one meter distance from a standard candle (or through one square foot at one foot distance). A standard candle emits 4π lumens (assuming it to emit equally in all directions) because there are 4π square meters on the surface of a sphere of one meter radius. A *lambert* is the unit of intensity per unit area (or "brightness") of an extended surface, whether self-luminous or lighted by an outside source. The brightness of a surface can also be measured in candle power per square centimeter. One candle power per square centimeter is equal to π lamberts. Why so many more or less similar units are used is a mystery into which we shall not attempt to penetrate.

Variation of illumination with distance. If a small source of light is thought of as being at the center of a hollow sphere, it will illuminate the inner surface of the sphere, and if the sphere is then doubled in diameter, the same light must be spread over four times the area, since the area of a sphere varies as the square of its diameter. Hence the illumination, i.e., the amount of light received on each square centimeter of surface, must vary inversely as the square of the distance from the source. If C be taken as the candle power of the source, I as the amount of illumination, and d as the distance, we have $I = C/d^2$. I is measured either in foot-candles or meter-candles.

This law does not hold exactly unless the source is small. In the extreme case when it is very large compared with the distance away from it, the illumination is independent of that distance. This condition would never be reached in practice, but one might approach it in certain cases. It is not safe to rely on the inverse square law when the width of the source is greater than one-tenth of the distance from it.

The photometer. The best way of comparing the candle power of two sources is by making them produce equal illumination on two adjacent parts of the same surface. This can be done very simply but not very accurately by setting up a vertical lead pencil in front of a white wall, and then placing the two lights so that the shadows of the pencil which they cast just touch each other on the wall and are equally illuminated; this arrangement is known as the Rumford photometer Fig. 30–1. Under these conditions the

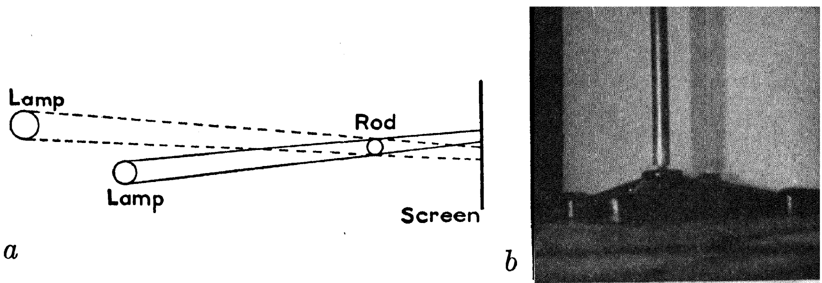


FIG. 30–1
The Rumford photometer

eye can make comparisons between the two with some precision. As, however, other objects in the room are likely to contribute light unequally to the two shadows, more accurate results can be obtained by somewhat more elaborate devices.

In any such arrangement when the illumination, I_1, on the screen produced by one light is equal to that, I_2, produced by the other, we have $I_1 = I_2$, or $C_1/d_1^2 = C_2/d_2^2$. Hence $C_1/C_2 = d_1^2/d_2^2$, which amounts to saying that the candle powers are directly proportional to the squares of the distances from each lamp to the screen. This is the equation of the photometer.

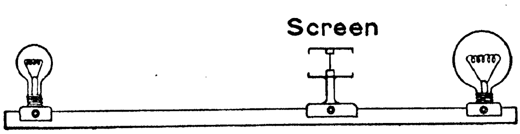


FIG. 30–2
A Lummer-Brodhun photometer

The best form of apparatus for this purpose is the Lummer-Brodhun photometer. In this arrangement the two lights are at the ends of a long graduated bar (Fig. 30–2) with a white plaster screen somewhere between them. The rest of the apparatus is merely a scheme for viewing with one eye parts of both sides of this screen simultaneously, and for varying the distances from the lamps to the

screen. Figure 30-3 shows that light which has come from one source is scattered from one side of the white screen, goes first to a mirror M and then enters a peculiar cubical piece of glass, and from this passes through a small viewing telescope to the eye. The cube consists of two pieces in contact at the center only. Through this region of contact light passes freely from the left side into the eye. Any light coming from that side and not striking the cemented spot is reflected back out of the way. Light from the opposite side of the plaster screen is reflected into the viewing telescope from all parts of the cube excepting the cemented portion. Hence the eye sees a field in the telescope, as in Fig. 30-4, the central spot being lighted from the left side of the plaster screen, the outer portion from the right. A very sharp line divides the two, and on account of this fact the eye is able to distinguish accurately when the two regions are equally bright.

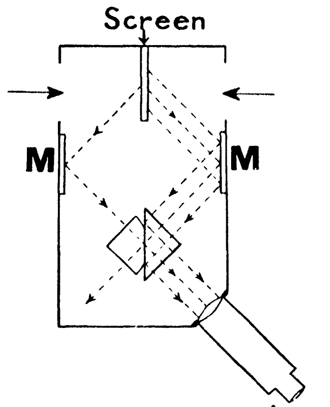


FIG. 30-3
The "head" of the Lummer-Brodhun photometer

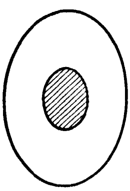


FIG. 30-4

With this instrument a precision of 1% is easy, and a skilled observer may do nearly ten times better.

Color effects in the photometer. When two surfaces of not quite the same color are brought close together, there is an effect in the eye which enhances their difference in color. This circumstance often makes trouble in photometric experiments. The hotter a source is, the whiter (or one might almost say bluer) is its light. Thus if the light of a common candle is compared in this way with the kind of light one gets from a gas-filled tungsten lamp, reduced to the same intensity, the color difference is so exaggerated that the candle light seems to be orange-yellow and the other bluish-white; and no exact "match" can be made in the photometer.

Flicker photometer. When the color difference between two lights makes it impossible to tell whether they are of equal intensity or not, another device may be used for making photometric tests. The eye has the peculiarity of continuing to give the sensation of light for a short but measurable time (about 0.1 sec.) after the light is cut off. This so-called "persistence of vision" is made use of in the *flicker photometer*. In this instrument the eye sees first one light, then the other, alternating at such a rate that the colors mix by

persistence of vision, and a mixture color is the only one that is seen. But if the rate of flickering is not too rapid, one sees this mixture color flickering in intensity (without alteration in color) except when the illumination from the two sources is the same. The accuracy attainable is not very high.

Practical illumination requirements. The person who has to plan the lighting of a hall or room needs to know first how many foot-candles of illumination should be supplied. As the result of experience, we now know that 20 to 30 foot-candles are desirable for close work like drafting, engraving, fine machine work, etc. For ordinary reading five foot-candles are good. A newspaper ***can*** be read with one foot-candle, but it is a strain on the eyes to do so. Very dull daylight gives 100 foot-candles; the full moon gives about 0.02; the sun at the zenith, 9600; starlight, 0.00008. Efficiency experts have discovered that more and better work is done if the illumination is rather more than the figures recommended, provided that glare is prevented.

The effect of glare is interesting. The eye has the ability to change its sensitiveness over an extremely wide range (see p. 508). If a very bright light strikes it, the sensitiveness of the whole eye reduces itself to meet the circumstances, and objects at one side seem dim, though actually brightly illuminated. When the direct glare is shaded off, the brightness of the rest of the room seems to increase, due to the ensuing gain in eye sensitiveness. Glare from a shiny page acts in somewhat the same way and makes reading more difficult. Thus it comes about that the amount of light recommended for proper illumination will not always prove large enough, if the lights are strongly concentrated and glaring.

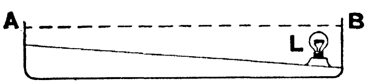


FIG. 30–5
Section of a foot-candle-meter

A portable ***foot-candle-meter*** (Fig. 30–6) is now available, which architects and others interested may carry about with them, enabling them to test the illumination in any part of a room and on surfaces held at different angles. In this compact instrument a box, containing a small flash-light lamp of a certain definite candle power, is covered with an opaque white card, AB, perforated regularly with holes as shown in Fig. 30–5, each of which is covered with thin translucent paper. The translucent holes near the lamp are brighter than the surface of the card, illuminated from the outside by the light under examination, while those near the end A

are dim and may appear dark in comparison with the white surface. If the lamp is always run at a constant voltage (as shown by a voltmeter which is part of the instrument) a scale may be marked on the card so that the place on the scale where the small holes seem to be of the ***same*** brightness as the adjacent surface will give the illumination directly in foot-candles.

Recent improvements in light sources. Our commonest sources of "artificial" light contain a piece of metal (usually a wire) heated to a very high temperature. The energy given off in radiation from such hot bodies consists almost entirely of invisible ("heat") rays. For instance, the newest and best tungsten lamp, filled with mercury vapor, furnishes only 11% of its energy in the form of light, a candle very much less. The firefly is usually pointed to as a model of economy which we must strive to imitate. It gives

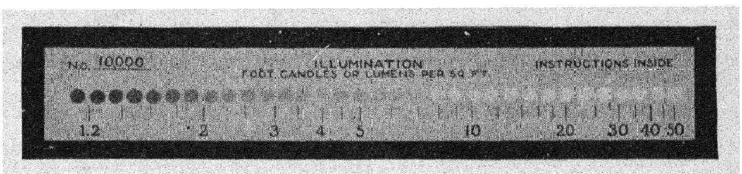


FIG. 30–6

The face of the foot-candle-meter, showing an illumination of about 7 foot-candles

"cold" light, that is, most of its emitted energy is in the form of visible rays. We have as yet no means of attaining any such efficiency.

Ordinary tungsten lamps give about one candle power per watt of power used (1 1/3 in large sizes to 4/5 in small lamps). Such sources as the mercury arc lamp are better in this respect, but have color peculiarities. Carbon arcs also are more efficient, but require an appreciable expenditure for maintenance. The improvements of the last twenty years have given us at least three times as much light for the same expense.

PROBLEMS

1. Considering the illumination and the effect of glare, what are the relative advantages of high and low supports for street arc lamps?

2. Why are the best lighted streets now supplied with many lamps of moderate brightness, rather than a few very bright ones? Why have these lamps diffusing shades?

3. What are the main points to be considered in providing a room with good illumination? Consider more than one type of room.

4. What is the illumination under a street arc lamp of 1500 candle power at a distance of 30 feet from the lamp? Could one read a newspaper there?

5. A reading table is to be supplied with an illumination of 2 ft.-candles from a lamp 8 ft. above it. How powerful a lamp will be required? How will this result be affected if the ceiling is a good reflector?

6. A photometer has a standard 20 candle-power lamp at one end and an unknown lamp at the other. The screen shows equal illumination from the two lamps when it is 3 ft. from the standard and 5 ft. from the unknown. Find the candle power of the unknown and the illumination on the screen.

7. Make a drawing of a photometer whose total length (from the standard lamp to the one being tested) is 2 meters. If the standard is just half as bright as the one being tested, find the proper position at which the photometer screen in the diagram should be placed.

THE NATURE OF LIGHT

Theories of light. In the early history of optical theories we find that light was supposed to be due to small ***corpuscles*** of an unknown nature. Plato believed that these were emitted by the eye, and produced the sensation of light when they struck the body seen, or something emanating from that body. Later the corpuscles were thought to be emitted by the luminous body and reflected into the eye. This theory was developed by Newton, whose authority kept it in the ascendant for nearly a hundred years after his death (1727), in spite of the fact that more and more experiments were being made whose results could not be explained except by the ***wave theory.*** In the twentieth century further discoveries have been made, especially under the heading of photoelectricity (p. 598), which, in their turn, are inconsistent with the wave theory, but require a corpuscular explanation. It is now evident that some combination of the two theories is necessary, based on an association of particles with waves, but the new Newton who will be able to present us with a fusion of these two rather contradictory points of view into one simple and consistent theory seems not yet to have appeared.

We shall proceed to consider first those facts that cannot be explained without assuming that light is due to a wave motion.

Simple diffraction effects. An early objection to the wave theory of light was based on the "fact" that light travels in straight

lines, while other waves do not. Water waves, or sound waves, bend rather freely around corners, while light was not known to do this until Fresnel (1818) and others set up experiments under proper conditions. The reason why this remained so long undiscovered was that the effects are on a very small scale, and unless the source is a very narrow one, its various parts give patterns

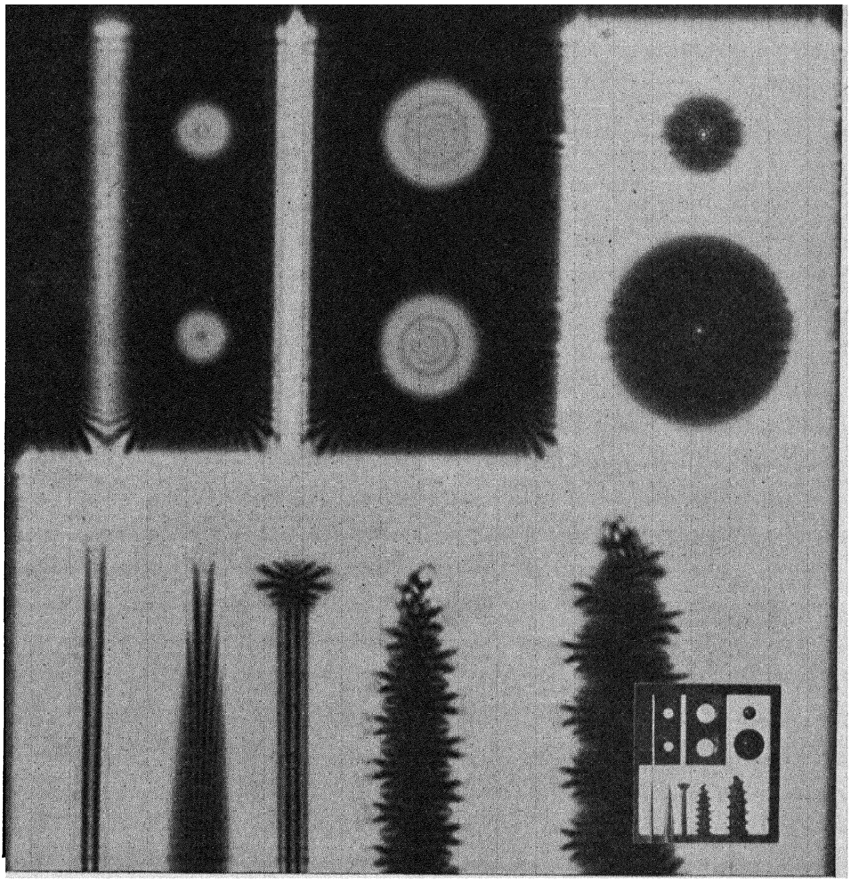


FIG. 30–7

which overlap, and hence nothing of interest is seen. Such experiments are best carried out by concentrating the light of an electric arc upon a small hole (say 0.2 mm.) in a metal plate, and making observations on the shadows of objects placed at distances of a meter or two from the hole. We then see amazing shadow patterns whose aspects change with the distance beyond the objects casting the shadows. Figure 30–7 shows an enlarged photograph of the

shadow about a foot behind an object illuminated in this way. The object itself is shown life-size in the lower right-hand corner of the same figure. It is a collection containing two slits, four circular holes of different sizes, two unequal round discs supported on very thin wires, two needles, a pin and two wood screws; incidentally there are corners and edges. The spreading of light around the edges of the slits, corners, needles, etc., produces fringes spreading into the shadows. These are so marked in the case of the wood screws that their sharp ridges appear to be entirely eaten away. In the holes there are rings, and all edges give fringes on the illuminated side. The explanation of all these figures on the wave theory is not difficult, but would delay us too much. Certainly no one after seeing these ***diffraction effects*** could ever again think that light travels in straight lines. The original patterns are more beautiful than any photograph, as they are colored.[1]

Diffraction around a round obstacle. The case of a circular disc, or a sphere, shown in Fig. 30–7, around which light waves from a point source spread, is historically famous. When Fresnel[2] was advancing the wave theory of light both by theory and experiment, the mathematicians, Poisson and Arago, calculated on this basis that one should find a bright spot in the very center of the shadow of a round object. This was regarded as an absurdity which disproved the theory until the experiment was carried out under proper conditions and the spot was found as predicted. We can see by very simple reasoning why this spot appears. The light waves, passing by the edge of the disc, spread somewhat into the shadow. To reach the central spot in the shadow, the waves must travel equal distances, no matter what part of the edge of the sphere they come from; hence all the waves arrive at this spot in the same phase and unite in a common effect. Any other point in the shadow is unsymmetrically placed, and while many waves reach it in the same phase, many others coming from different parts of the surface of the sphere arrive in opposite phases, so that

[1] On account of photographic difficulties, Fig. 30–7 fails to show clearly the rings inside the holes, or the fringes produced by the slits. If a different time of exposure had been chosen other features would have been lost.

[2] A. J. Fresnel, (1788–1827), French physicist and engineer. In addition to his work on diffraction his name is associated with apparatus which he invented for the study of interference and polarization. He gave strong support to the wave theory of light by his theoretical and experimental investigations, which admirably supplemented each other.

the net result is a complete destruction of the light by interference (see p. 561) and there remains only the central white spot, surrounded when the object is small by a set of faint rings.

We shall return to a consideration of diffraction phenomena later (p. 550). At present it is interesting to note merely how obvious they make the spreading of light around corners, and the reader may be willing to assume for the present that all the observed phenomena of this sort can be explained on the basis of the wave theory of light.

Huygens' principle. Huygens was one of the early supporters of the wave theory of light, though not quite in its modern form. He has left us an important principle which helps us to see how diffraction and other effects arise.

When we drop a stone into water, a series of waves in the form of rings grows with a center at the spot where the stone falls. By the time the rings are a few feet wide, the stone has gone, and the growth of the rings thereafter cannot logically be traced to any action on the part of the stone. It must be that the waves at any one instant generate those existing at the next. Huygens gave this a general form by saying that in any sort of wave ***each particle in the wave-front may be regarded as an independent source of disturbance,*** generating the wave arrangement that exists in the next instant by combining its effects with those produced by all the other particles in the wave-front, as shown in Fig. 30–8. This idea will serve us later in explaining several phenomena, when it will be seen to yield in a very simple manner results in harmony with the facts. No disturbances are found anywhere except along the new wave-front because the little waves ("wavelets") from the individual sources destroy one another at such places by mutual interference. The motion still existing in the part of the medium over which the wave has just passed can be shown to destroy any backward-moving wave that might otherwise be expected to occur.

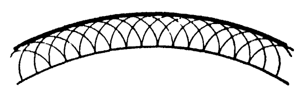


FIG. 30–8

THE SPEED OF LIGHT

Early measurements. The problem of determining the speed of light is a famous one, which has occupied the attention of many experimenters from the time of Galileo to the present. It is

attractive both on account of its difficulty and because of its theoretical interest, since the electromagnetic theory predicts that the speed of light should be equal to the ratio of the electrical units (p. 366). Even without any theory to guide us we might still infer from our common experiences with waves that light should have a definite speed if it has a wave nature; while, if it is formed of corpuscles, these like bullets might well be very variable.

The first successful attempt was made by Roemer, a Danish astronomer (1675), from observations made on the revolution of the moons of the planet Jupiter. Laboratory methods were devised by Fizeau, a French physicist (1848), and after him by a long series of experimenters, each one introducing improvements. One of these was Michelson,[1] who spent two years (1880–1882) on the problem, and has quite recently returned to it again. He has apparently succeeded in attaining a precision of at least 1 part in 100,000. His latest method, being the best of all, deserves some attention.

Michelson's experiment, (1926–1929). A distance of about 23 miles was surveyed with extraordinary care and skill by the U. S. Coast and Geodetic Survey, from Mt. Wilson to Mt. San Antonio in California. The result was 35,426.3 meters, and this was considered accurate to within 0.1 meter. Over this distance a beam of light was sent from Mt. Wilson and returned by a mirror on Mt. San Antonio, and the time taken on this journey was determined. This is the most difficult part of the task, as the journey occupies only about 1/9000 second. To measure so short a time within one part in 100,000 (i.e. almost to one-thousand-millionth of a second) was the feat which Michelson accomplished.

The vital part of the apparatus was an eight-sided mirror of steel or fused quartz which could be rotated at a high rate of speed and maintained at this rate with extraordinary accuracy and constancy. The light beam struck one face (*a*) of this mirror when this was at rest (Fig. 30–9), was reflected off on its long journey, and on its return, following almost the same path back, struck

[1] A. A. Michelson, Professor of Physics at the University of Chicago; winner of the Nobel prize in physics in 1907; ingenious and accurate experimenter whose work has been concentrated mainly in the field of optics; inventor of the interferometer and the echelon grating; author of two admirable books, concerned largely with his own researches, "Light Waves and their Uses," (University of Chicago Press, 1903), and "Studies in Optics," 1927.

another face (c) and then entered the observing telescope. When the mirror was started rotating the light beam found the face (c) at a new angle on its return, which increased more and more as the rotation grew more rapid. Finally such a speed was attained that face (b) reached the position formerly occupied by (c) when the light returned. That is, during the time required for the long journey, the mirror turned through one-eighth of a complete rotation. The optical device (telescope and cross-hair) provided to test the exactness of the return path from the second mirror was very precise, and the method of measuring the rate of rotation of the mirror was equally so. The eight-sided mirror in one experiment ran at the rate of 528.76 revolutions per second, and therefore took one-eighth of 1/528.76 sec. to turn one face around to the plane of the next, which was the time the light took to travel twice 35.4263 km.; whence the speed of light was 299,711 km. per second. This experiment being done in air yields a value when reduced to vacuum of 299,796 ($\pm$ 4) km. per second or 186,285 miles per second. Successive measurements agreed to within a few kilometers. Figure 30–9 shows the general arrangement of the apparatus, considerably simplified, and not drawn to scale.

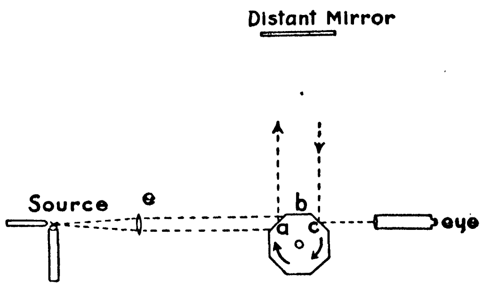


FIG. 30–9
Michelson's experiment on the velocity of light

Speed of light in other media than air. In a vacuum light travels a very little faster than it does in air; in water about three-quarters as fast; in ordinary glass about two-thirds as fast. These facts have been established by laboratory experiments in tubes, (first by Foucault in 1850), which we shall not take the time to describe. It is, however, important that light is slowed down by matter when it has to pass through it. On the corpuscular theory, a beam of light, striking slantwise on a smooth surface of water and being bent downwards on entering the water (as it is), could do so only by an attraction between the corpuscles and the water, which would inevitably result in speeding up the corpuscles as they entered it. The discovery of the fact that this does not occur was a death blow to the corpuscular theory.

PROBLEMS

1. How many times would light go around the earth at the equator in one second, if the earth's diameter is 8000 miles? (Radio waves are actually observed to do this.)

2. If light vibrates 20,000 times in going a distance of one centimeter, what is its frequency?

3. If a room were lined with mirrors reflecting 96% of the light falling on them, and a light which had been burning there were suddenly obliterated, why would there not be a prolongation of light in the room analogous to the reverberation of sound?

CHAPTER 31

REFLECTION AND REFRACTION OF LIGHT

Reflection. We are all well acquainted with the reflection of light from a mirror, from a surface of still water, or from a polished metal surface. We know that a transparent substance can reflect light, though some opaque ones (e.g. polished silver) do it much better. The ***amount*** of light reflected depends on the angle at which the light strikes the surface. Anyone who has watched a clear sunset over a large body of still water may have noticed the brightness of the light reflected when the sun is near the horizon. Though the sun is not then so brilliant as usual, on account of atmospheric absorption, the reflected light is nearly as bright as as the direct sunlight. The ***direction*** in which light is reflected from a surface is an important problem which is easily attacked by the use of Huygens' principle.

Direction of reflection. Huygens' principle shows us in what directions light must be reflected from a smooth surface. Let AB be the surface, and AC the front of a wave advancing upon this surface from a great distance, so that the front in the diagram (which is part of a circle with the source as center) is practically a straight line. The point A is represented as just being disturbed. The point F will be disturbed as soon as the light has had time to travel over the distance CF. In this time the wave is thrown back from A, and that point forms a wavelet whose radius is equal to CF. A point G midway between A and F will have time to send a return wavelet only half as far, and this is accordingly drawn in Fig. 31–1 as a dotted circle of radius equal to half of CF. Intermediate points between A and F will generate wavelets of proportionate sizes, one of which is drawn in the figure. The net effect of all these is a complicated crossing and intermingling of wavelets, with mutual

and complete destructive interference at all points except along FH, the common tangent to all the wavelets. FH is thus the new wave-front, and AH the new direction, in which the reflected wave moves. The symmetry of the figure shows that DA and AH must be equally inclined to the line NA, which is the "normal" (the perpendicular to the surface). Hence the angle DAN usually called the ***angle of incidence,*** is equal to the angle NAH, called the ***angle of reflection.*** This is the law of reflection from plane surfaces, the truth of which can readily be tested in the laboratory.

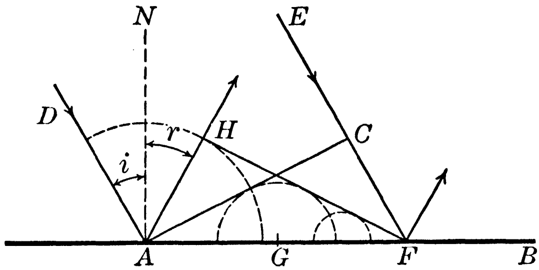


FIG. 31–1

Images in plane mirrors. As a consequence of the law of reflection, it is easy to show that a small light will send its rays out to a mirror in all directions, but these will be received by an observer (of course on the same side of the mirror) as though they had come from a "virtual image" of the source as far behind the mirror (along a perpendicular line) as the real source is in front. If S (Fig. 31–2) is the source and LM the mirror, the rays between SL and SM are the only ones striking the mirror and therefore the image I can be seen only within the angle FIH. The geometry of this figure is in accord with the law of reflection. The proof of the exact location of I is left to the reader.

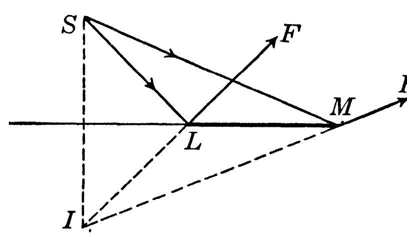


FIG. 31–2

Reflection in general. Reflection occurs whenever light strikes a surface of separation of two media in which it has a different speed. For instance, when light enters a plate-glass window about 4% of it is reflected at the front surface of the glass, and the remaining 96% proceeds on through the glass to the rear surface. At this second surface the light is again reflected in the same proportion; this is spoken of as a reflection in glass against air, and reminds one of the reflection that occurs in sound (p. 253) when a wave traveling along inside a

pipe comes to the open end. A plate of glass thus gives two reflections, returning about 8% of the light in ordinary circumstances. Objects thus seen by reflection may appear doubled especially if the two surfaces of the glass are not parallel. The same effect is often observed in the case of a glass mirror, silvered on the back, where the front reflection is relatively faint, but may cause trouble.

It is reflection that makes transparent objects visible. A glass rod reflects light coming from the windows or lights in the room. If the end of it is wet with Canada balsam (or, better, chloral hydrate dissolved with heat in a little glycerine), in which light has nearly the same speed as in glass, the liquid becomes part of the solid, optically speaking, and the glass itself seems to be dripping off in liquid form from its own tip. A clean cut-glass bowl will become invisible if placed in a uniformly lighted enclosure, such as a box painted on the inside with luminous ("phosphorescent") paint.

Reflection occurs from rough surfaces, accompanied by scattering in all directions. A piece of glazed paper scatters somewhat,

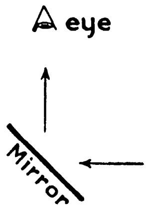


FIG. 31–3
Reflection

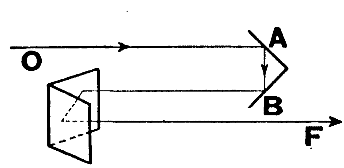


FIG. 31–4
Inversion by reflection

but also reflects regularly, so that it sends more light in the direction given by the law of reflection than in any other.

Inversion on reflection. A single reflection inverts objects in one direction. Thus, a man's right hand appears to be his left, when seen in a mirror. A landscape appears upside down if we view it as in Fig. 31–3 in a mirror held at 45° below the eye. Four reflections can be so arranged that the light passes on in the original direction, but inversion occurs in ***both*** directions. Figure 31–4 shows two mirrors, A and B, perpendicular to each other, and a beam of light, coming from O, being reflected from A to B and thence to a second similar pair whose line of junction is perpendicular to that of the first, so that the second reflections occur in a horizontal plane, if the first are vertical. The final ray F is parallel to the original ray O, but slightly displaced sideways.

Such a device is used in prism binoculars to invert an image without the use of any lenses.

Refraction. When light strikes a transparent surface, a part of it enters and travels through the second material, taking a new direction unless it strikes perpendicularly. Let us suppose that a wave-front AC comes in air (as Fig. 31–5) from a distance and strikes a plane surface, AB, say of glass. While the light along EC is traveling from C to F in air, that along DA is partly returned as a reflected ray (which we examined on p. 482, but shall now ignore), but the main part of it goes forward. Its speed in ordinary glass is about two-thirds that in air. Hence the Huygens wavelet starting downward from A will make a circle on the diagram whose radius AH is two-thirds of CF. Intermediate points between A and F will form smaller wavelets in proportion (as in the reflection case on p. 482) and the net result of the combination of all these is a new wave-front in the direction HF, and a new direction of the ray, AH. If NP is the normal to the surface, the angle DAN is, as before, the angle of incidence, and PAH is the angle of refraction, which we shall call b (because the ray is bent). In the triangle CAF the angle CAF is equal to i, since it and the angle DAN each make a right angle when they are added to the angle NAC. The sine of the angle $CAF = \sin i = CF/AF$. Likewise $\sin b = AH/AF$. Hence

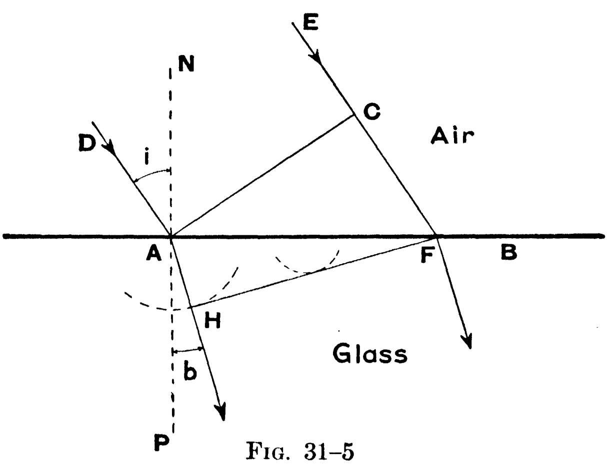


FIG. 31–5
Refraction

$$\frac{\sin i}{\sin b} = \frac{CF}{AH} = \frac{\text{distance traveled by light in air}}{\text{distance by light in glass in the same time}}$$
$$= \frac{\text{speed of light in air}}{\text{speed of light in glass}} = n.$$

The ratio of these speeds is called the ***index of refraction*** of the glass and is denoted by n. By this simple reasoning we have ar-

rived at Snell's law of refraction (1621), $n = \sin i/\sin b$. It is interesting to note that Kepler and other early workers tried for a long time to discover this relation, but in vain. It is often true that the difficulties of one scientific generation seem absurdly easy to the next.

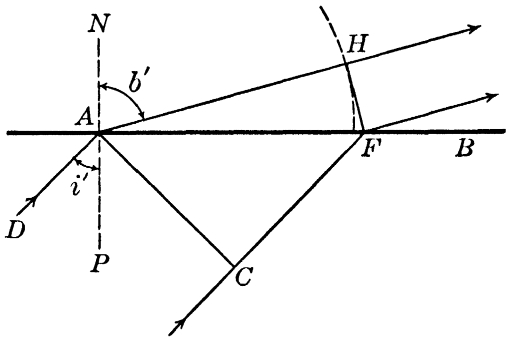


FIG. 31–6
Refraction from glass to air

Refraction as from glass to air. If a Huygens diagram is made for the case of a plane wavefront AC advancing in glass to a plane surface bounded by air, as in Fig. 31–6, we see that the wavelet in air with center A must be drawn with radius $AH = \frac{3}{2}CF$, since it has all the time to travel in air that the wave in glass takes to cover the distance CF. Hence the incident ray DA is bent to AH, this time ***away*** from the normal NP. The angle DAP, though in glass, should now be called the angle of incidence, i', and the angle NAH becomes the angle of refraction, b'. As before $n = \sin b'/\sin i'$. The relation seems inverted this time, on account of inverting the direction of motion of the light and the names of the angles.

Total reflection. In the case just considered an interesting effect occurs at a slightly larger angle of incidence, Fig. 31–7. The wavelet in the air with center A may now have a radius equal to AF, and it and all other wavelets will then pass through F simultaneously. At any angle greater than this the wavelets will all pass beyond F, and no tangent can be drawn to them from F, which means physically that they destroy one another by mutual interference in the upper medium and no wave occurs there, the light being ***totally reflected*** back into the glass. The limiting

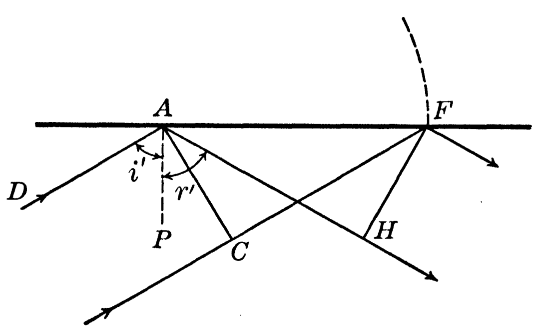


FIG. 31–7
Total reflection

value of i' at which this happens is when $n \times CF = AF$ or when $\sin b' = 1$. Hence $n = 1/\sin i'$, if i' is now understood to have this limiting value. It is then called the ***critical angle*** of incidence, and the last equation shows that if it can be measured, as it readily can, the index of refraction of the material (or the speed of light in it) is immediately obtained. Instruments for doing this are called ***refractometers;*** they are much used in optical laboratories. It is not necessary to use more than a very small specimen of the material (or a small drop, if liquid) in order to carry out the measurements.

Total reflection is easily shown as a lecture experiment by sending the light from a small arc (Fig. 31-8) through a lens to make a parallel beam and directing it by a mirror M upward through the wall of a small rectangular glass tank full of water colored with a little uranin, or other fluorescent dye. The track of the light in the water then shows beautifully by a greenish glow in the yellowish solution, and if the tank is full to the brim, the ray escaping (if any) may also be traced by receiving it on a screen S. Changing the angle of the mirror M will then cause the incident angle of the ray under water to pass through its critical value, and the changes in intensity of the reflected and refracted beams are very marked. For small angles of incidence the reflected ray is feeble and the refracted strong. Making the angle of incidence change through the critical angle reveals a sudden change in the intensity of the reflected ray, as it then becomes as strong as the incident. The critical angle for water is about 49°, and for ordinary glass about 41°.

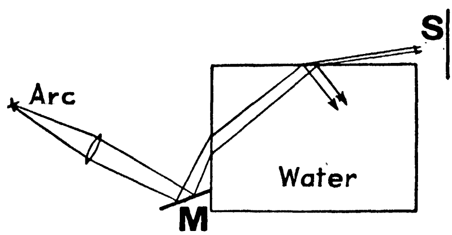


FIG. 31-8

The refraction of light must enable a fish looking up toward the surface of the water to see opposite banks of the lake in which it is swimming, not at their usual angle of 180° apart, but at an angle which is twice the critical angle, i.e., about 98°. Beyond this angle total reflection sets in and there appears an inverted view of the bottom, if there is light enough. Naturally this experiment succeeds in still water only.

A drinking glass, full of water, shows silvery reflections when the surface is looked at from below at the proper angle.

Experiment illustrating reflection and refraction. An interesting mechanical model can easily be made to imitate the phenomena of reflection and refraction. It consists simply of an axle 3 or 4 inches long carrying two independent wheels, 1 inch in diameter, which enable it to roll with little friction along the surface of a wide board. It travels in a straight line unless it meets a "medium" in which its velocity is different. In Fig. 31–9 it is shown approaching a cloth-covered surface where friction will hold back wheel A while wheel B is still on the clear board, and "refraction" will occur, bending the direction of travel over toward the normal to the surface. In case Fig. 31–10*a* the wheel is in the cloth-covered portion approaching the medium in which its velocity is greater, at an angle of incidence less than the critical angle. When wheel A gets out, it gains on B, and the roller is "refracted" away from the normal. In case Fig. 31–10*b* the angle of incidence is greater than the critical angle, and wheel B never gets out, as A gains so much that it turns the whole roller back into the slower medium. The roller is "totally reflected."

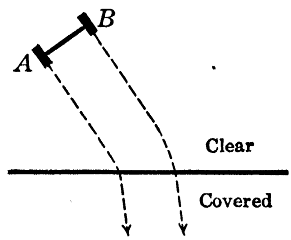


FIG. 31–9

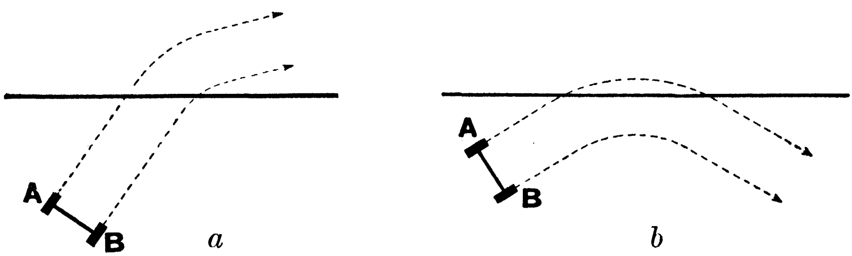


FIG. 31–10

Applications. Total reflection is used in optics to produce a mirror which reflects light without serious loss and ***gives only one image.*** A totally reflecting 45° prism of glass, cut as shown in Fig. 31–11, allows a beam of light to enter normally at A, be totally reflected at B and emerge at C with no losses except the unimportant ones due to reflection at A and C (p. 482). Naturally the three surfaces must be optically good, or they will introduce defects in images formed by the instrument of which they are a part (e.g., a field glass, or a periscope). Sometimes, however, the accuracy of these surfaces is unimportant. For instance, rows of such prisms are used

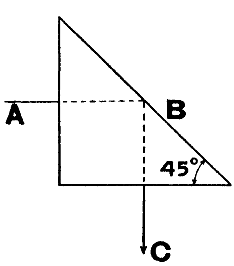


FIG. 31–11
A totally-reflecting prism

among tall buildings for the purpose of catching vertical light and throwing it horizontally into dark spaces, cellars, etc.

Curiously enough a totally reflecting prism works perfectly only when the reflecting surface is very clean. Dirt on the surface scatters some of the light and appears from the back side as though it were brightly illuminated. A protective coating of varnish or paint would spoil the prism as a reflector. A small hole in such a coating would allow a single Huygens wavelet to escape. The complete destruction of the light behind this surface and therefore the usefulness of the surface as a reflector depends paradoxically on all the wavelets getting through so that they may destroy one another by mutual interference and return their energy to the glass to go into the reflected beam.

A diamond is much more brilliant than a similar piece of glass. The difference lies in the fact that the diamond has an index of refraction of 2.42 and a critical angle of about 25°. Thus the chance that a ray of light will be totally reflected inside the jewel and returned in the direction of our eyes is much greater than it would be in the case of glass.

Mirages. If the air is stratified in temperature (or moisture content) the speed of light through it varies with the height above the ground level. Thus over a tarred or cement road heated on a sunny day, the speed of light is greater near the surface than above, and a ray going down toward it at a grazing angle may be bent in a curve and turn upwards again. A person receiving such a ray considers it to have been reflected from a water surface. The "pools" of water which he sees on the road dry up magically when they are approached because the angle at which they are seen becomes too great.

The converse happens over cold bodies of water, where distant shores, icebergs, etc. "loom" up to magical heights. The rays are then curved so as to be concave to the earth's surface. The same effects happen with sound waves (p. 249).

PROBLEMS

1. What is the index of refraction of a substance whose critical angle is 30°?

2. Find the velocity of light in a diamond, if its index of refraction of 2.4.

3. Prove from the law of reflection that your image in a mirror is as far behind it as you are in front.

4. What is the smallest mirror in which a man can see himself from head to foot, and how must it be placed?

5. Show that a ray of light reflected from a rotating mirror turns in a given time through an angle which is twice as great as that through which the mirror turns in the same time.

6. If a distant light is viewed by reflection at 45° in a thick plate of red glass, two images are seen, one white, the other red. Explain.

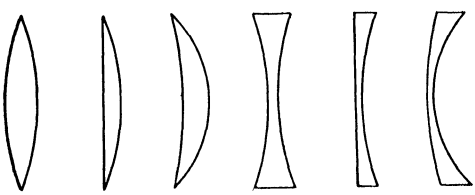
7. A passenger on a train looking as far ahead as he can see through the window sees a street arc lamp accompanied by several faint images of itself which appear to approach the lamp as the train comes to be more nearly opposite to the lamp. Explain.

CHAPTER 32

LENSES AND CURVED REFLECTORS

Lenses, 490; action of lenses, real images, 490; virtual images, 491; finding images by rays, sizes of images, 492; proof of the lens formulæ, 493; the lens-maker's equation, 495; the image equation, 496; lens powers, diopters, 497; concave mirrors, 497; thick lenses, 499; focal length of a thick lens, or of a group of lenses, 499; the aperture of a lens, 500; spherical aberration, 501; astigmatism, 502; chromatic aberration, 503.

Lenses. Lenses are transparent bodies with curved surfaces whose uses have been known from ancient times. They were known as burning glasses to the Greeks, used in spectacles in the thirteenth century, and in such combinations as telescopes and microscopes in the beginning of the seventeenth. Their function is to receive rays of light and so bend them that they pass through a point of concentration (a focus), and there form an image of the source from which they came. Strictly speaking, a lens is a transparent body bounded by spherical surfaces which intersect, or nearly so. In Fig. 32–1 are shown in section the common shapes of lenses, (*a*) being double-convex, (*b*) plano-convex, (*c*) meniscus, (*d*) double-concave, (*e*) plano-concave, and (*f*) convexo-concave. The first three are thicker in the middle than at the edge, and will be shown to be of the type called ***converging*** lenses; the opposite is true of the last three, which are ***diverging*** lenses.



a *b* *c* *d* *e* *f*

FIG. 32–1

Action of lenses. Real images. A wave-front which passes through a lens has its curvature changed because light travels more slowly in glass than it does in air, and hence the times of passage through the center of the lens and through its edge are different.

If the source of light is small and is a long distance away, the

wave-front is a portion of a large sphere, so that a small piece of it, such as PQ, Fig. 32–2, will be a portion of a plane. If this passes through a lens which is thicker at the center than the edge, the central part of the wave-front will be delayed, and the front will come out curved, as in $P'Q'$. Light in air can be regarded as moving along directions, called ***rays,*** which are at right angles to the wave-front. Hence such a wave-front as $P'Q'$ will shrink down to a point I, where these rays meet. This forms an image of the distant source, and a piece of paper placed there will catch it and make it easily seen. Any image that can thus be caught is called a ***real image;*** i.e., a real image is formed where a bundle of rays of light converge, and cross.

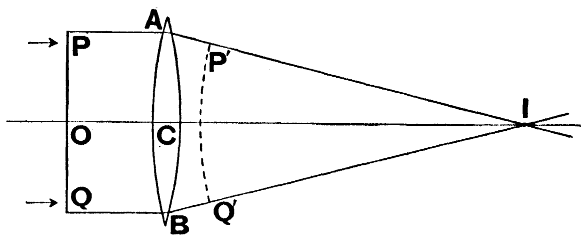


FIG. 32–2

In such circumstances the distance CI from the lens (which is assumed to be thin) to the image is known as the principal focal length, or more often simply the ***focal length*** of the lens; that is, the focal length, is the ***distance from the lens to the image of a distant object.*** If the object should approach the lens, the wave-front PQ would become part of a smaller sphere, that is, it would become more curved (O bulging out toward C), and then the action of the lens, delaying the central rays, would be unable to produce so much backward curvature in $P'Q'$, with the result that the image would be farther away than before. Thus, as the object comes in toward the lens, the image recedes from it. At one place they are equidistant; then when the object comes still closer and reaches a distance equal to the focal length, the wave-front $P'Q'$ has become plane and the image is then at an infinite distance.

Virtual images. If the object continues on its journey toward the lens and reaches a point O inside the focal length CF (Fig. 32–3) the wave-front PQ is now so strongly curved that the lens cannot reverse or even annul its curvature, so that it emerges as $P'Q'$, curved in the same direction as before. The rays then diverge on the right-hand side of the lens, and no image is formed there. But $P'Q'$ is now a portion of a sphere with center at some point I' on

the same side of the lens as the source. This means that if the wave-front $P'Q'$ entered the eye of an observer, he would interpret it as having come from an object at I', or, in other words, he would "see" an image at I'. Such an image is like the one seen in a mirror in that it is not a spot where rays of light meet, but merely one from which they appear to have come. We call these ***virtual images.***

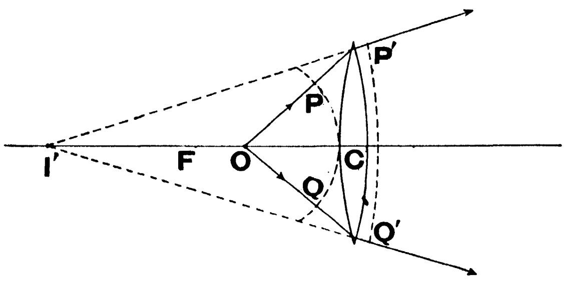


Fig. 32-3

We get virtual images from diverging lenses also. If we place the object back at a great distance again, and use a lens which is thicker at the edge than at the middle, we shall get a wave-front $P'Q'$ of the sort just discussed (Fig. 32-4) coming from a virtual image I' somewhere on the left side of the lens. The rays will diverge beyond the lens, following lines drawn perpendicular to $P'Q'$, which are, of course, radii of the sphere whose center is I'.

A divergent lens is incapable of forming real images by itself. Such lenses are used in combination with strongly convergent lenses, which overpower them, so to speak, so that the combination is convergent.

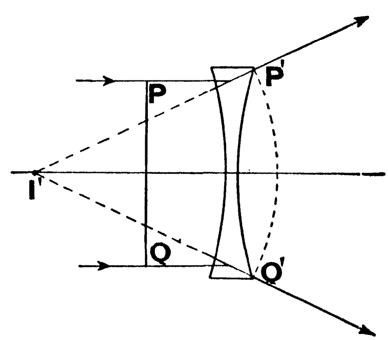


Fig. 32-4

Finding images by rays. Sizes of images. Strictly speaking it is not correct to regard light as traveling along straight lines, or rays, but in many cases the idea is useful in leading to accurate results in a simpler way than by following the motions of wave-fronts. An example is furnished by the graphical method of finding the position of an image formed by a lens. The axis of the lens is OCI (Fig. 32-5), drawn in the direction perpendicular to the lens at its middle point. The object OA is supposed to be at any distance from the lens. The focal points F are to be marked on both sides of the lens. A ray AL is then drawn from a point on the object, parallel to the axis. Since all

lines parallel to the axis are like rays coming from a very distant object and converge beyond the lens at its focal point F, this particular ray must do so, and will then follow the course LF. A second ray can be drawn from the point A whose course we can easily trace; this is the one through the center of the lens. All rays go through the center of a thin lens without change of direction because that part of the lens is nothing but a thin plate with parallel sides. Hence the ray AC must pass on beyond and will intersect LF at the point D. Since two rays have thus been shown to start from A and meet at D, there must be an image of A at D and all the rays leaving A and reaching the lens must meet again there. Thus the position of the image is determined.

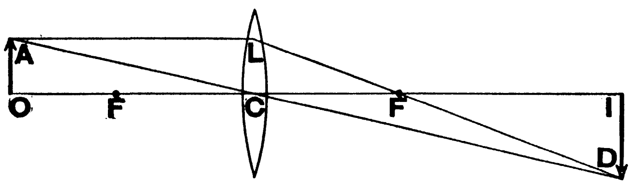


FIG. 32–5

The image is real and inverted under the conditions shown in the figure. The ***size of the image*** evidently depends on how far away from the lens it is formed. The similar triangles AOC and DIC yield the ratio

$$\frac{\text{size of image}}{\text{size of object}} = \frac{\text{distance from lens to image}}{\text{distance from lens to object}}$$

which will be found useful.

If the lens is a diverging one, or if the object is so close that the rays diverge beyond the lens, the lines LF and AC do not meet on that side, but they will then meet, if produced backward, on the ***same*** side of the lens as the object. Their intersection again gives the true position of the image, which is now a virtual one and erect.

Proof of the lens formulæ. It is so important in connection with optical instruments to be able to calculate the position of an image formed by a lens that we must find the formula needed for that purpose. The derivation of this formula follows simply from the ideas of wave-fronts already considered.

A wave is supposed to start from the object O and generate a forward-moving wave-front, which at one instant takes the form ACB, just touching the lens L (Fig. 32–6). This wave then ad-

vances, and at the moment when it emerges from the lens it may take the form DEF, the middle of the wave having been delayed by its passage through the glass. This wave-front will then travel

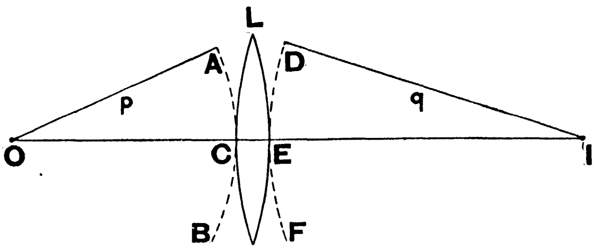


FIG. 32–6

forward shrinking as it goes and form an image at I, the center of the circle DEF. The lens is supposed to be small and thin, so that the distance of A or of D from the lens is negligible compared with the long distance p from the object to the lens. The time the light takes in passing from A to D through the air (the edge of the lens is supposed to be infinitely thin) must be equal to the time taken in passing through the thickness CE in glass. Thus

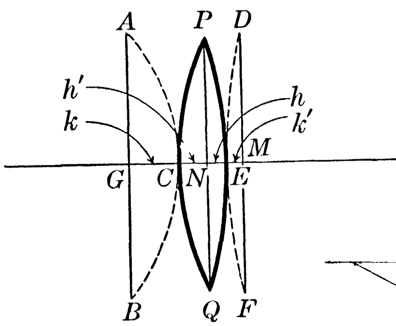


FIG. 32–7

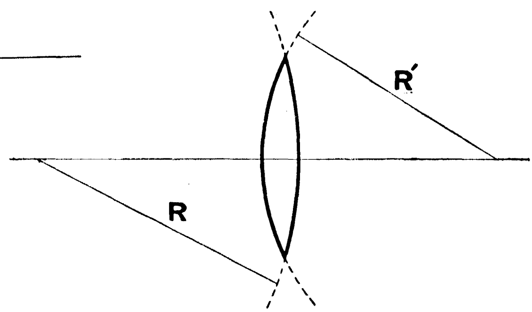


FIG. 32–8

$$\frac{AD}{CE} = \frac{\text{velocity in air}}{\text{velocity in glass}} = n, \text{ the index of refraction of the glass, or}$$

$AD = n \times CE$. In Fig. 32–7 ACB is the approaching wave-front, and DEF the escaping one. The short distance GC, cut off on the axis between the chord AB and the arc ACB will be called k; similarly, $CN = h'$; $NE = h$; and $EM = k'$. Thus the distance $AD = GM = h + h' + k + k'$; while $CE = h + h'$.

Therefore the equation just stated yields

$$h + h' + k + k' = n(h + h'), \text{ or}$$
$$(n - 1)(h + h') = k + k'$$

But, by a well-known geometrical proposition[1] $k = d^2/8p$, where d is the chord, which is here the diameter of the lens. Similarly $k' = \frac{d^2}{8q}$, $h = \frac{d^2}{8R}$ and $h' = \frac{d^2}{8R'}$, where R and R' are the radii of curvatures of the lens surfaces (Fig. 32–8). Hence

$$(n-1)\left(\frac{d^2}{8R} + \frac{d^2}{8R'}\right) = \frac{d^2}{8p} + \frac{d^2}{8q} \quad \text{or} \quad (n-1)\left(\frac{1}{R} + \frac{1}{R'}\right) = \frac{1}{p} + \frac{1}{q}.$$

The left-hand side of this equation is constant for a particular lens, since it refers only to the dimensions and the index of refraction. We can see what this constant must be by noting that $1/p + 1/q$ must be equal to the same constant. We have already noted that if the object is at an infinite distance the image is at the focal distance, f. In such a case $1/p$ is zero, and $1/q = 1/f$; but, if $1/p + 1/q$ is equal to $1/f$ in this case, and is constant, it must always have this value. Hence

$$\frac{1}{p} + \frac{1}{q} = \frac{1}{f} \text{ and } (n - 1)\left(\frac{1}{R} + \frac{1}{R'}\right) = \frac{1}{f}.$$

These two equations that have thus been reached are worth considering in detail.

The lens-maker's equation. The second of these equations may be called the lens-maker's equation because it tells him how he must grind the surfaces of a lens in order to produce a required focal length, the index of refraction of the glass being known. It

[1] In the circle in Fig. 32–9 the chord BD cuts off a length FA from the radius CA. The chord is assumed to be short compared with the radius. In the triangle BCF,

$$BC^2 = CF^2 + FB^2 = (CA - AF)^2 + FB^2$$
$$= CA^2 + AF^2 - 2CA \times AF + FB^2.$$

But $BC = CA$, and AF^2 is negligibly small; hence

$$2\,CA \times AF = BF^2,$$

or

$$AF = \frac{\text{(half the chord)}^2}{\text{twice the radius}} = \frac{\text{(the chord)}^2}{\text{8 times the radius}}.$$

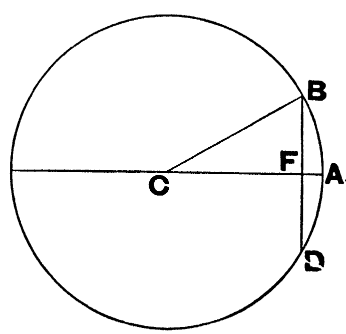


FIG. 32–9

does not, however, bind him to doing it in only one way, since there are two radii of curvature, R and R', and only one equation to be satisfied. He may choose one of these radii as he wishes, and then the equation fixes the other one.

For example, if one side of the lens is to be plane ($1/R' = 0$), and the focal length is to be 20 cm. (and $n = 1.5$), the equation becomes $(1.5 - 1)1/R = 1/20$, or, $R = 10$ cm.; that is, the curved surface must be part of a sphere of 10 cm. radius. If both surfaces are to be curved alike, $R = R'$, and the equation again takes a simple form. If the next batch of glass which the lens-maker buys proves to have a slightly different index of refraction, as it well may on account of differences in composition, he will need to use slightly different radii in order to make lenses of the same focal lengths.

One might wonder what use there is in making lenses of different shapes. The reason is that one can thereby reduce the defect known as spherical aberration (p. 501).

The image equation. The equation $1/p + 1/q = 1/f$ might be called the image equation, because it enables us to calculate the distance of the image (from the lens), the focal length and the object distance being known. The description of lens action already given on p. 491 is all condensed into this little equation. Thus, if the object is at a great distance $1/p = 0$, and $q = f$; that is, the image is at the focal distance from the lens. If the object approaches, $1/p$ increases, and therefore $1/q$ must decrease; or the image moves away. At one point they are at equal distances; then the image equation gives $2/p = 1/f$ or $p = 2f$. Each is at a distance $2f$ from the lens, and the object is at a distance of $4f$ from the image. From the relation between size and distance (p. 493) it follows that the image is then equal to the object in size. As the object comes closer to the lens, q must increase until it becomes infinite at the moment when $p = f$. If p becomes less than f, we have $1/q = 1/f - 1/p$, which is then negative. A negative value of q implies that the image is on the other side of the lens from where we had assumed it to be. Thus q now gives us the distance of the ***virtual image.***

The same formula holds for finding the position of the virtual image formed by a ***diverging lens,*** but we must agree to treat diverging lenses as having negative focal lengths. The focal length of a diverging lens is defined as the distance of that point

from which rays appear to have come which were parallel when they approached the lens. Thus Fig. 32–10 shows rays from a distant object approaching a double-concave lens and, after passing through it, diverging as though coming from the point F.

The image equation may also be used in some cases to find the focal length. For example, one might be asked to buy a camera lens of such a focal length that one could photograph diagrams with it on a scale of half their natural size, with the added condition that the distance from the lens to the photographic film (or plate) could be no greater than six inches. When the image is half the size of the object, it must be half as far away (p. 493) from the lens. Hence the object distance is 12 inches; $p = 6$ inches; and hence $1/f = 1/6 + 1/12$ or $f = 4$ inches. Other problems of this class are given below.

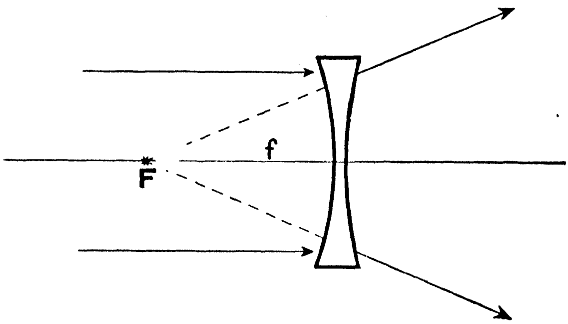


FIG. 32–10

Lens powers. Diopters. In opticians' language the *power* of a lens is proportional to the reciprocal of the focal length, because a short-focus lens gives larger images at moderate distances. It is usually measured in units called *diopters*. A lens of f centimeters focal length is said to have a power of $100/f$ diopters; that is, the number of diopters is the number of times the focal length goes into a length of one meter. A lens of 50 centimeters focal length has a power of 2 diopters; one of 25 centimeters focal length has 4. It is shown later (p. 512) that close combinations of lenses have powers which are very conveniently obtained by addition. Thus, a lens of 4 diopters and one of 2 when placed together are equivalent to one of 6. A converging lens of 4 diopters combined with a diverging lens of -2 would make a combination with a power of $+2$.

Concave mirrors. Concave mirrors form images whose distances are given by the same image equation as for lenses. Their action is, of course, quite different, being due to reflection rather than refraction. A spherical concave surface will reflect rays starting from the center of its sphere back upon themselves, and form an image coincident with the object (though inverted); and if the object recedes from the mirror, the image approaches it, until

when the object is very distant the image is found at a distance again called the focal length, and easily shown to be half the radius of the sphere. The image formed by a *small* spherical reflector may be of very good quality and is not affected by the common lens defect known as chromatic aberration (p. 503), as all colors are reflected at the same angle. If the spherical reflector is made larger, defects in the images due to spherical aberration (p. 501) enter in and spoil their sharpness. Better images of distant objects are obtained if the reflector is made parabolic rather than spherical. The difference in shape is very slight, but the parabolic surface has the property of reflecting all rays coming parallel to its axis accurately to the same point (Fig. 32–11). Thus if a parabolic reflector of very large size is made, it gathers up a great quantity of light from a distant source; more than the eye does, in the ratio of its area to that of the opening of the eye. Such a mirror 100 inches wide is used in the great Mt. Wilson reflecting telescope, which collects 150,000 times as much as the eye can and discloses stars much fainter and more distant than were ever known before. A 200-inch mirror is planned with which our knowledge of the remote parts of the universe will be much increased.

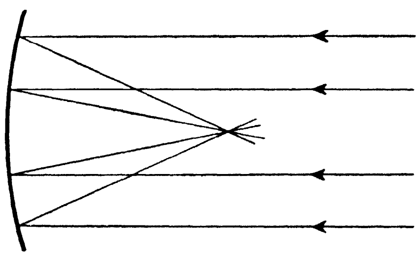


FIG. 32–11
A parabolic reflector

A parabolic concave reflector may be used to concentrate invisible "heat" rays from distant objects and make them detectable. Thus at night the radiation from a candle a mile off is easily measurable, and even a man or a cow may similarly be revealed by the radiation emitted. The heat from several of the brighter stars has also been studied with such apparatus, and the temperature of Mars and other planets determined.

Concave parabolic reflectors are much used in the opposite way, that is, by placing a small source of light at the focus and sending out from the reflector a parallel beam. This is done in automobile headlights (though here a certain amount of sideways scattering is purposely introduced), and in searchlights. The long, straight shaft of light sent by a searchlight across the sky is made visible to us by the scattering by fog or dust particles along its track, and shows us how very successfully the beam has been made parallel.

If it spreads slightly as it proceeds, this arises from the fact that the powerful arc light which is the real source is too large for all of it to lie accurately at the focus of the reflector.

Thick lenses. The image equation that is given above is derived on the assumption that the thickness of the lens considered is negligible in comparison with its focal length. While this is usually accurate enough, it is not satisfactory in work of the highest precision; also, very thick lenses are not uncommon, for which the approximation no longer holds. In such cases two points (P and Q, Fig. 32–12) known as the *principal points*, can be found which have useful properties. The object distance, formerly described as the distance from the object to the (negligibly thin) lens, should now be taken to the first principal point; and the image distance similarly taken from the second. If this is done, it can be shown that the image equation still holds. In a thin lens the principal points come together, and coincide with the center of the lens.

It can also be shown that a ray passing toward one of these points emerges along a parallel line drawn through the other point. Thus it travels through the central part of the lens with a little jog sideways on account of the thickness.

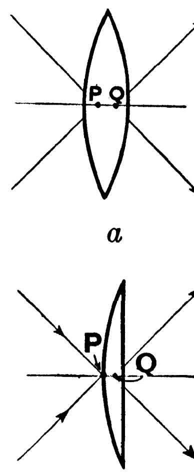


FIG. 32–12

If the *principal planes* are drawn (perpendicular to the axis through the principal points), any ray directed toward any point on one of these planes leaves the lens as though it had come from a point on the other which is equidistant from the axis. The second ray is, in general, not parallel to the first. A ray parallel to the axis, meeting one principal plane, leaves the corresponding point on the other plane and passes through the principal focal point. This rule enables one to find the position of images by the same graphical method already used on p. 493.

The calculation of the positions of the principal points, while not difficult, is omitted here.

Focal length of a thick lens, or of a group of lenses. The focal length of an ordinary thin lens is the distance from the image of a distant point to the lens; but if the lens is thick, and especially if it is unsymmetrical, the point to which the measurement must be taken becomes the nearer principal point. The same remark applies to combinations (close groups) of lenses. In such cases one sometimes hears the term *back focus*, meaning the distance from the back surface to the image. More commonly the term *equivalent focal length* is used, which is measured from the principal focus to the nearer principal point in the lens.

The focal length of a thick lens may easily be measured in any practical case. Let O be the object and I the image when the group of lenses is at L (Fig. 32–13). The object and image in any case are called *conjugate*, because they may be interchanged in position. Putting the object at a distance from

the lens equal to LI is equivalent to leaving it at O and moving the lens to L'. The image will then be found at I as before (though now it will be smaller). Thus if O and I are fixed (and far enough apart), two positions can be found

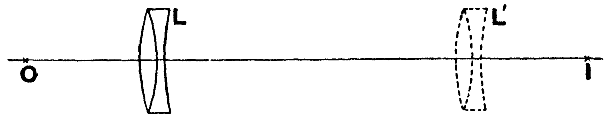


FIG. 32–13

for the lens which will give the image at I. The image equation then enables us to deduce a simple connection between LL' (called D) and OI (called L). Thus:

$$\begin{aligned} OL &= L'I = p. \\ LI &= OL' = q. \\ OI &= L = p + q. \\ LL' &= D = q - p. \end{aligned}$$

Adding the last two,

$$2q = L + D.$$

Subtracting,

$$2p = L - D.$$

Multiplying these,

$$4pq = L^2 - D^2.$$

Thus

$$\frac{1}{f} = \frac{1}{p} + \frac{1}{q} = \frac{(p+q)}{pq}$$

becomes

$$f = \frac{pq}{(p+q)} = \frac{L^2 - D^2}{4L}.$$

Hence the equivalent focal length f is determined by measuring the distance between the object and the image as well as the distance *through which the lens has to be moved,* which can be done without knowing anything about the lens itself.

The aperture of a lens. The ***relative aperture of a lens,*** (sometimes simply called the aperture) is the ratio of the diameter to the focal length. It is an indication of the amount of light which the lens gathers up to put into the image. For example, compare two lenses, each of 1 inch diameter, one of 10 inches focal length, the other of 20. The first forms an image of an object. The second forms an image of the same object which is twice as wide and twice as high, covering, therefore, four times the area. In a camera the latter lens would require four times as long an exposure of the same object, since the same amount of light is spread out (like butter on bread) so much thinner.

The usual way of stating the relative apertures of camera lenses is best shown by an example: a lens of 2 inches diameter and 12 inches focal length is said to have an aperture of $f/6$, i.e., its diameter is 1/6 of its focal length. A modern camera lens of $f/2$ aperture gathers up an amount of light in proportion to the square of its diameter, i.e., 16 times as much as an $f/8$ lens; and takes good pictures in one-sixteenth of the time. The ***numerical aperture*** ("n.a.") of microscope objectives is defined as the ratio of the ***radius*** of the lens opening to the distance from the object to the edge of the lens (multiplied in the case of oil-immersion lenses by the index of refraction of the oil).

Spherical aberration. It is a property of lenses with spherical surfaces that rays from a distant object are not all brought to a single focal point unless the lens has a small numerical aperture. Figure 32–14 shows this defect greatly exaggerated. Rays striking the lens near the edge are bent to a nearer focus (B) than those which pass near the center (and meet at A). The importance of this defect depends on the distance AB, as well as on the focal length and the numerical aperture of the lens. The distance AB can be shown to depend on the shape of the lens and to be proportional to the focal length and the square of the numerical aperture. For a simple lens of 10 cm. diameter and 100 cm. focal length (numerical aperture 1/10) the distance AB is 1.67 cm. if the lens is a symmetrical double-convex one; 4.5 cm. if the lens is plano-convex with the plane side toward the distant source, and 1.17 cm. if the plane side is turned toward the image.

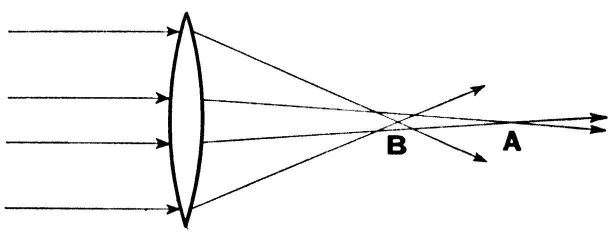


FIG. 32–14
Spherical aberration

It is possible to avoid, or at least reduce, this trouble in three ways. The ***first*** is to reduce the diameter of the lens, or to "stop it down" by placing a diaphragm in the beam near the lens so as to cut out all but the central rays. This reduces the amount of light transmitted by the lens, which is generally a bad thing to do. The ***second*** method is to reduce the amount of spherical aberration to a harmlessly small amount by choosing a lens of the best shape,

as indicated above, or better, by combining lenses of different shapes. Thus it is possible to combine a strongly convergent lens of a shape producing a small amount of aberration with a weaker diverging lens which gives the same aberration in the opposite direction without at the same time completely annulling the convergence of the rays produced by the first lens. There is no reason for considering this method in detail, as spherical aberration is never treated by itself. Actually, chromatic aberration (p. 503) also enters in, and can be eliminated by a suitable choice of lenses in combination. The combinations that are chosen are such as to eliminate both types of aberration at once. The ***third*** method is to abandon spherical surfaces and grind the lens to a slightly different shape. This has usually to be done by hand, and the expense of the process confines its use mainly to the manufacture of large astronomical telescope lenses, where the utmost precision must be obtained regardless of cost. "Aspherical" (i.e. not spherical), condensing lenses are deservedly coming into use in the better sorts of projection lanterns.

Astigmatism. When rays go through a simple, thin lens at a considerable angle to the axis (Fig. 32–15), the images which are

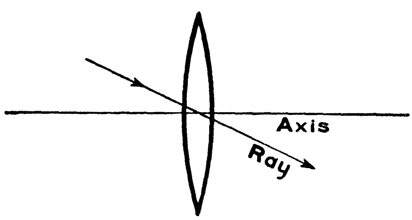


FIG. 32–15

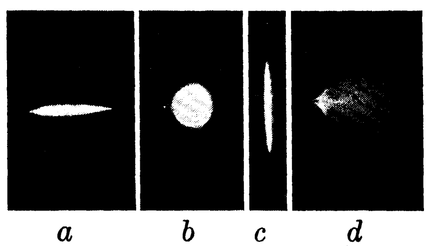


FIG. 32–16
Images of a point source

formed are not good. A point source gives rise to a line image, Fig. 32–16*a* or (at a different distance) a line at right angles to the first Fig. 32–16*c*. Between these a circular patch Fig. 32–16*b* may be found, which is the nearest approach to an image of the point source. The formation of line images of this sort is called ***astigmatism.*** It is exhibited still more clearly in the images formed by concave mirrors at an angle to the axis. When a lens of fairly large aperture is used to form an image of a point source at an angle, the effect of astigmatism is modified (by the addition of spherical aberration) into a figure called "coma," which can take several forms,

one of which is shown in Fig. 32–16*d*. Evidently an uncorrected lens in a camera cannot be expected to give very good images except in the direction of its axis. The corners of the picture are commonly not very sharp when taken with a cheap lens.

Chromatic Aberration. The bending of light by a lens varies with the color. A simple lens, if it had no other defects, would bring the blue rays in white light (p. 524) to a nearer focus (*B*, Fig. 32–17) than the red (*R*). The yellow rays, which affect the eye most, come in between. A screen placed so as to catch the yellow focus would show an image with a purple border, the outer edges being bluish.

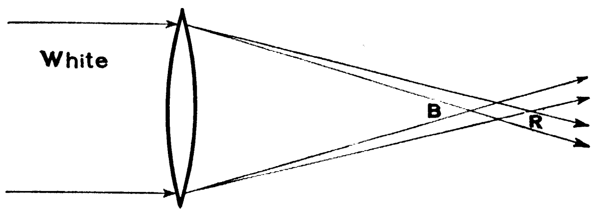


FIG. 32–17
Chromatic aberration

The correction of this defect is accomplished by combinations of lenses made of different sorts of glass. It is explained below, (p. 525).

PROBLEMS

1. Use the geometrical method of p. 493 to find the image formed by a diverging lens ($f = -5$ cm.) of an object which is 4 cm. from this lens.

2. Find the position of the image of an object which is 9 cm. from a lens whose surfaces have equal radii of curvature of 10 cm., and whose index of refraction is 1.5.

3. Find the index of refraction of the glass of which a lens is made, if its two surfaces have each a radius of curvature of 12 cm., and the lens has a focal length of 10 cm.

4. A lantern slide, 3 in. wide is to be projected on a screen at a distance of 20 ft. by means of a lens of 6 in. focal length. How wide a screen will be needed in order to catch the whole picture?

5. A lens-maker has a lot of glass whose index of refraction is 1.6. He wishes to make a plano-convex lens of 20 cm. focal length. What curvature must he give to the convex surface?

6. An unsymmetrical double-convex lens has two surfaces whose radii of curvature are 50 and 60 cm. respectively. If its focal length is 50 cm., what must be the index of refraction of the glass of which it is made?

7. A camera with a lens of 6 in. focal length is fitted with a long bellows. It is desired to photograph a 12-in. drawing with this camera, so as to have the picture 3 in. long. How far from the lens must the drawing be?

8. Going back to the proof on p. 494, show that a lens of index n' immersed in a liquid of index n has a focal length given by

$$\frac{1}{f} = \left[\frac{(n' - n)}{n}\right]\left(\frac{1}{R} + \frac{1}{R'}\right).$$

9. Using the formula of the preceding problem, find the focal length in water (index 1.33) of a glass lens (index 1.5) whose focal length in air is 12 mm.

10. A small hand camera is made with a lens of opening $f/2$. If its focal length is 4 in., how wide is the lens itself? How much of this lens does one use if he "stops it down" to $f/8$?

11. Plot a curve with p and q as coördinates, showing what geometrical form the image equation has.

CHAPTER 33

OPTICAL INSTRUMENTS

The uses of optical instruments. The ingenuity of man is nowhere more evident than in the various ways in which he has enlarged his natural powers of vision by means of optical instruments. Lenses have been found in the ruins of ancient Egypt, where were also jewels so delicately cut as to require a magnifier in the hands of the craftsman who produced them. In the thirteenth century the use of spectacles for the correction of faults of vision was not uncommon, and the earliest crude forms of microscope and telescope, such as one might make by combining two spectacle lenses, may have been known then to Roger Bacon. In Galileo's time (1608) the first telescopes of which there are any records were produced, and one of his own devising first revealed to him some of the secrets of the heavens with such astonishing results that it was regarded by many with superstitious fear. The microscope began at the same time to disclose a world of living things of undreamed-of smallness and has since led through medical research to a great increase in the span of human life. Besides the camera, the motion-picture projector and other common optical instruments, we now possess others whose functions are purely scientific, such as the spectroscope and the interferometer. Many of these instruments are worth examining in detail. Before doing so it would be well to consider first the action of the earliest optical instrument employed by man, i.e., his own eye.

The eye. The eye is a comparatively simple optical system. The front spherical surface consists of a transparent shell C, the

cornea, directly behind which comes a small chamber A, filled with a weak salt solution, the aqueous humor. Next comes the crystalline lens L, built up of layers of tissue of varying index of refraction, the most refracting being in the center. The beam of light entering this lens is limited by the iris, which varies in width of opening from over 7 to less than 3 millimeters with increasing brightness of the light. Beyond the crystalline lens is the main body of the eyeball, filled with a thin jelly called the vitreous humor, at the back of which is the sensitive surface, the retina, upon which the image falls. The volume A is really a lens, L is another and V constitutes a third at the limit of which the final image is formed. Thus the eye consists actually of a group of lenses. Most of the refraction occurs at the front surface, but this is added to by the action of the crystalline lens which has an average index of refraction (1.44) greater than that of any other part of the eye.

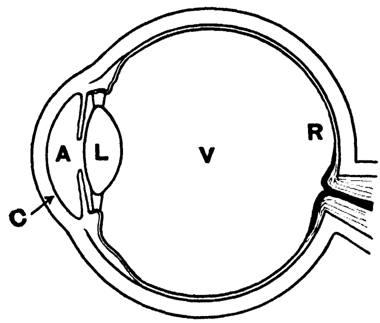


Fig. 33–1

Accommodation. The back surface of the crystalline lens can be curved by means of muscles attached to it. If its curvature is increased, the converging power of the crystalline lens is increased, and rays of light diverging from a near object can be brought to meet again on the retina. A flatter rear surface is appropriate for viewing distant objects. Thus in the eye the image distance remains constant, and changes in the object distance are compensated for by changes in focal length. If the focal length were constant, the front of the eye would have to move out or in to accommodate it for near and far objects. This arrangement would look odd in a human eye, but is actually found among the fishes.

Astigmatism, and other defects. If the lens surfaces in the eye are not quite spherical, they act like combinations of spherical and cylindrical lenses. A cylindrical lens gives a line image of a point source similar to that formed by an ordinary lens (p. 502) when the rays pass through it at an angle. Thus people whose eyes suffer from this very common defect are able to see some of the lines radiating from a common center, as in Fig. 33–2, more distinctly than others. In the direction in which the lines are sharp each point on the line

makes a short line image parallel to the line itself, and thus produces a clear image of the whole line. In a direction at right angles these images make the lines fuzzy. The axis of the cylinder in the eye is parallel to the distinct lines in the diagram. The eye also suffers from the defects known as spherical and chromatic aberration. The first of these seems to be partially eliminated by the iris diaphragm which cuts off some of the most aberrant rays (at least when the pupil is small), and by the construction of the crystalline lens L, which is not uniform, but is more strongly refracting toward the center than near the edge. The chromatic aberration is ordinarily very little noticed, but seems to be uncompensated. It becomes very marked when bright objects are viewed through a cobalt-blue glass, which cuts out all the rays to which the eye is most sensitive, and allows the deep red and blue to pass.

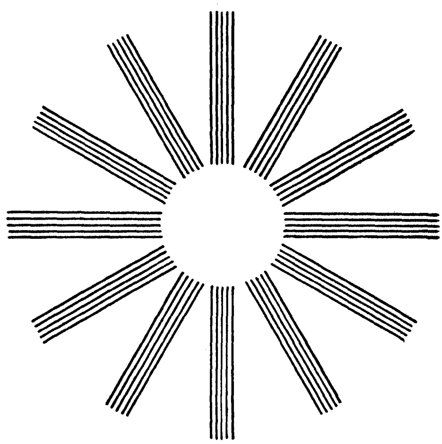


FIG. 33–2
A test of astigmatism

The retina. The sensitive surface on which the image is formed is called the ***retina.*** The light falls first upon a layer of nerves, and passing through this surface reaches the nerve endings beyond, which are of two sorts — rods and cones. The sensation of sight appears to be generated in these. The cones are most concentrated in the center of the retina, where one's vision is best, and they appear to be associated particularly with the process of color detection. The rods are more numerous elsewhere, and are believed to be more sensitive to light than the cones, but to be incapable of indicating differences of color. The eyes of nocturnal animals are furnished entirely with rods; the catfish has two sets, rods for daylight and cones for night work, which can be thrust forward alternately. Many nocturnal birds and animals have a reflecting layer behind, presumably to help in concentrating the light on the cones; such eyes "glow in the dark"; that is, they can reflect light which is furnished by an artificial source, such as a flash light or an automobile headlight. They are not self-luminous.

If the light becomes very bright, the iris contracts, thus reducing the amount of light entering the eye by a factor of about ten times (proportional to the area of the opening). But the illumination out of doors on a sunny day is millions of times as strong as it is

on a starlit night, when we can see well enough to walk about, after our eyes have become accustomed to the darkness. Thus the eye needs further protection against strong light. In some animals (e.g., the rat) this is furnished by a layer of a black substance which partially covers the sensitive nerve endings in bright light and retires out of the way in the dark. In our eyes a different device is used. The sensitiveness is controlled by the amount of a substance developed in the nerve endings themselves which is called the ***visual purple.*** The light breaks up this substance (bleaches it) and this act seems to be the cause of the nerve impulses which produce vision. This colored substance absorbs most strongly those colors to which the eye is most sensitive. Though the visual purple is continually being produced, the amount of it in existence at any moment is very small in a strong light; thus the light can break up very little, and therefore the eye is not then very sensitive. In the dark the color has time to accumulate, and a feeble light can then be absorbed, and thus "seen." On entering a dimly lighted room from bright sunlight, we need a time of ten minutes or so before we can see clearly, and during this time the supply of visual purple is being replenished. In this simple way the sensitiveness of the eye can be changed by a factor of about a million, to keep pace with the changes in the illumination.

It is interesting to note that a strong light anywhere in the field of vision spreads so much inside the eye as to reduce its sensitiveness considerably. Thus a glaring light in a room with darkish walls is unpleasant, and we can really see better under such conditions if the light is covered with a large shade, even though the total amount of light in the room is thereby reduced. The gain in sensitiveness of the eye more than makes up for the loss of light. This is one of the chief advantages of the system of indirect lighting.

Vision. The nerve impulses in the eye have been proved to have electrical features, and it may be that the act of seeing is, after all, mainly an electrical phenomenon. Nerve impulses travel at much more moderate speeds (three meters a second) than electrical signals along wires, but they are complicated, and the study of such matters is beyond the scope of this book.

When a lightning flash lights up the landscape, the act of vision persists for a certain time (about one-twentieth of a second), even though the flash may last only a thousandth of this time, or less.

This effect is known as the persistence of vision. It can be measured by observing a light, which can be made to flicker by sending it through a perforated disc revolving at variable speed. At high speeds the light appears perfectly steady; at low speeds it flickers. At the critical speed when the flickering just disappears, the time interval between flashes gives the time of persistence of the sensation. This time is found to vary with the color. In viewing motion pictures the persistence of vision is one of the chief aids in securing the illusion of motion, and it makes us unaware of the gaps between successive pictures.

The fact that we have two eyes, which see but one image of an object, indicates that the two retinal images are connected with only one part of the brain. These images are not, however, identical. We see slightly farther around the right side of an object with the right eye than with the left. This produces the

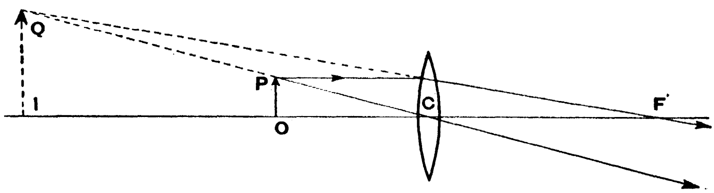


FIG. 33–3

stereoscopic effect by means of which we are enabled to estimate the distances of objects and see them in relief.

The simple lens as magnifier. A converging lens of short focal length yields enlarged virtual images of small objects, if we look into the lens, and hold the objects closer to the lens than the principal focus. The formation of these images has already been described, and the positions and sizes of the images can be found by methods given above (p. 496). If a lens of one-inch focal length is used and the object is placed one inch from the lens, the image which is then seen by an eye close behind the lens is as large as possible, and the rays entering the eye are parallel, as though the object were a long way off.

The geometrical construction for locating images (p. 493) enables us to find an image IQ (Fig. 33–3) of an object OP which is within the focal distance. Evidently, Q may be pushed a great distance away by making OC nearly equal to the focal length. The figure shows us that $\frac{IQ}{OP} = \frac{IC}{OC} = \frac{q}{p}$, and the image equation yields by

simple algebra $\frac{q}{p} = \frac{f}{p - q}$. Hence $\frac{\text{size of image}}{\text{size of object}} = \frac{f}{p - f}$; or, the actual size of the image varies greatly with the position of OP in Fig. 33–3. This implies that the image can easily be made infinitely large, but it is to be noted that when it is large, it is also far away. The important thing about a virtual image is not how large it is, but how large it appears to be, and this depends on the angle which it subtends at the lens; that is, on the ratio IQ/IC. Now this ratio is the same as OP/OC, and is practically constant for all distances actually used. Hence it might seem that we gain nothing in looking at small objects with a "magnifying" lens. What we really gain is the ability to see them clearly when they are much too close to the eye for this to be possible without the lens. The average eye can see objects sharply up to about 10 inches (25 cm.) from the eye, a distance generally known as ***the distance of most distinct vision.*** For shorter distances a certain amount of eyestrain is produced by the effort required to focus on the object; and no normal eye can see objects clearly within 5 or 6 inches, no matter how it is strained. But the 1-inch "magnifier" enables one to see an object clearly and comfortably at 1 inch, and then it looks 10 times as large as it did at 10 inches. Hence we say that a 1-inch lens magnifies 10 times when used in this way, assuming 10 inches (or 25 cm.) as the standard distance for such comparisons.

It is often stated that the object should be placed at such a position that the image is at the distance of most distinct vision. If a man had to dig a sliver out of his finger with a needle, he would certainly place his finger at that distance; but there is no advantage in placing a virtual image there because (as explained above) it looks no larger there than at a very great distance. Moreover, there is a distinct disadvantage, for the reason that the distance of most distinct vision is ***not*** the ***distance of most comfortable vision.*** There is a little eyestrain (increasing with age) associated with prolonged use of a short distance. The instinctive and most comfortable arrangement is to place the image at a considerable distance, so that the rays entering the eye are practically parallel, as they are from ordinary objects.

If the rays between the lens and the eye are parallel, it makes no difference whether the eye is close to the lens or a few inches behind. It is obvious that the virtual image will not look perceptibly smaller if its distance, which is already large, is increased

by a small amount. It is best to place the eye close to the lens, however, because the extent of the ***field of view*** is then increased, and this advantage is usually worth securing.

The simple magnifier appears, though usually in not quite so simple a form, as the ***eyepiece*** in such instruments as telescopes and microscopes. Since eyepieces are so common it is important that this account of their action should be clearly understood.

Magnifying power. As explained above, a 1-inch lens magnifies 10 times, if we take 10 inches as the standard distance at which to hold the unmagnified object. One may test this by placing a short piece of a millimeter scale under the lens in front of the right eye, and with the left eye viewing another millimeter scale 10 inches away. By a certain amount of mental jugglery, which becomes quite easy after a little practice, the two images may be superimposed and compared in size.

If a lens of f inches focal length is used for such a comparison, its magnifying power is easily seen to be $10/f$, or $25/f$ if f is expressed in centimeters.

Combinations of lenses. When two lenses are used together as a combination, they form images at distances which can easily be

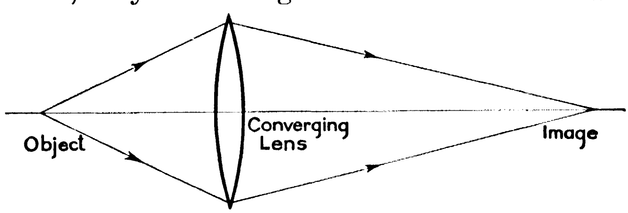


FIG. 33–4

found. One has only to ***calculate the position of the image formed by the first lens, as though the second one were not there; then, using the distance of this image from the second lens as though it were now the object, find the position of its image as formed by the second lens.*** The object distance in the second case will have to be considered as negative whenever the image formed by the first lens is beyond the second one, and therefore on the "wrong" side. Figure 33–4 shows the conventional arrangement of object, lens, and image, the light always being supposed to go from left to right. If the object, or the image, is on the other side (compared with its position in this diagram), it must be taken as negative. If the lens is a diverging one, it has a negative focal length. Examples will make this rule clear.

Examples of lens combinations. The reader is urged to make drawings for each of the following cases, indicating the positions of object and image.

(1) Consider first a thin converging lens of 12 cm. focal length and a thin diverging lens of -20 cm. focal length placed so close together as to be practically coincident. A distant object produces an image 12 cm. from the first lens and therefore 12 cm. (on the wrong side) from the second lens also. Hence the object distance for the second lens is -12 cm., and the image equation for it becomes

$$-\frac{1}{12}+\frac{1}{q}=-\frac{1}{20}$$

whence $q = 30$ cm., or the focal length of the combination is 30 cm.

Using powers in diopters (p. 497) the positive lens has a power of

$$\frac{100}{12}=8.333;$$

the second lens has a power of -5. Adding the powers, the sum is 3.333 and the focal length of the combination is

$$\frac{100}{3.333}=30 \text{ cm.}$$

(2) A positive lens of 6 cm. focal length is combined with a second one of 4 cm. focal length with a distance of 1 cm. between the lenses. A distant object has an image formed by the first lens which is 6 cm. from it, or 5 cm. from the second, again on the wrong side. Hence for the second lens

$$-\frac{1}{5}+\frac{1}{q}=\frac{1}{4} \quad \text{or} \quad q = 2.22 \text{ cm.}$$

measured from the center of the second lens.

In such a case the problem cannot be solved by adding the powers of the lenses on account of their separation.

(3) An object is 10 cm. from the first lens of the combination just considered. Where is the final image? The first image distance is given by

$$\frac{1}{10}+\frac{1}{q}=\frac{1}{6}, \quad \text{or} \quad q = 15 \text{ cm.}$$

This is 14 cm. from the second lens, on the wrong side. Hence for it

$$-\frac{1}{14}+\frac{1}{q}=\frac{1}{4} \quad \text{and} \quad q = 3.11 \text{ cm.},$$

which is the distance of the final image from the second lens.

(4) The first lens has a focal length of 50 in., the second of 2, and there are 60 in. between the lenses. Find the final image of an object 400 in. from the first lens. For the first lens

$$\frac{1}{400}+\frac{1}{q}=\frac{1}{50}, \quad \text{or} \quad q=\frac{400}{7}=57.14 \text{ in.}$$

For the second lens,

$$p = 60 - 57.14 = 2.86 \text{ and } \frac{1}{2.86} + \frac{1}{q} = \frac{1}{2};$$

whence $q = 6.65$ inches beyond the second lens.

(5) If the distance of separation of the lenses in the last problem had been 59 in., p for the second lens would have been $+ 1.86$, and q would have come out $- 26.6$ in., the sign indicating that the final image was a virtual one 26.6 in. from the second lens or 32.4 in. from the first; that is, lying between the lenses.

The telescope. The earliest type of telescope, used by Galileo and by the contemporary instrument-makers in Holland, is treated below (p. 517) under the heading of the opera glass. It has only a limited usefulness because of its narrow field of view. The simple form now used chiefly by astronomers consists essentially of two lenses. One of these is a converging lens of fairly long focal length,

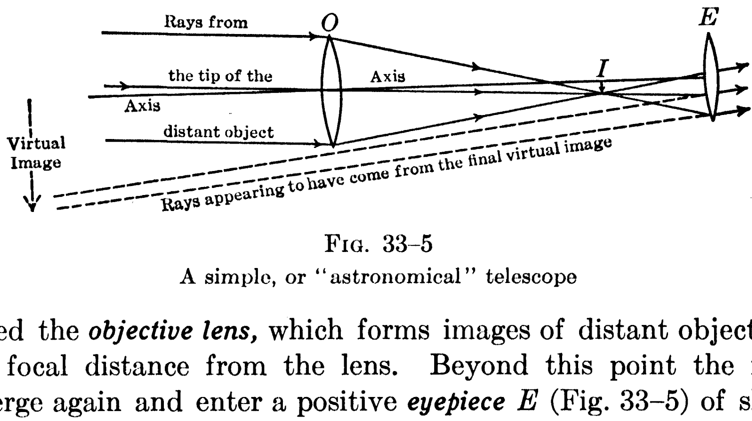


FIG. 33–5
A simple, or "astronomical" telescope

called the ***objective lens,*** which forms images of distant objects at the focal distance from the lens. Beyond this point the rays diverge again and enter a positive ***eyepiece*** E (Fig. 33–5) of short focal length. The rays emerge parallel, or nearly so, entering the eye as though they had come from a distance comfortable for prolonged observation. It is to be noted that all the light reaching O comes out through E. Thus the telescope has a great light-gathering effect and enables us to see faint objects.

The image I inside the telescope tube is real and inverted, and the eyepiece, acting like a simple magnifier (p. 509), leaves it inverted. For astronomical purposes this makes little difference, but for observations of objects on the surface of the earth we need a ***terrestrial type of telescope,*** in which the final image is erect. In one form (Fig. 33–6) this result is secured by the addition of the erecting lens L which gathers the rays diverging beyond the first image

I_1 and converges them to a real and erect image I_2 (usually a little magnified also) which is viewed through the eyepiece E as before. This form of telescope is evidently longer, and it usually follows on

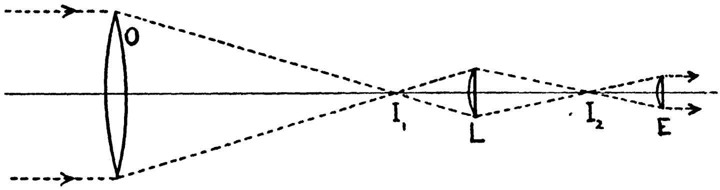


FIG. 33–6
A terrestrial telescope

this account that its field of view is narrower. The addition of the erecting lens also adds to the likelihood of optical defects in the image.

A better though more expensive way of inverting the image is by reflecting the rays inside the tube by means of totally reflecting prisms whose action is similar to that of the mirrors already described (p. 483). By this device the rays are folded upon themselves and the tube greatly shortened. Also, the inversion of the image takes place without introducing any optical defects, if the surfaces of the prisms are plane, and the field of view remains as large as it is with the astronomical type of telescope. Figure 33–7 shows the arrangement of the prisms and lenses in a ***prism binocular.***

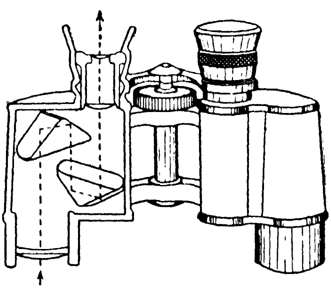


FIG. 33–7
A prism binocular

The exit pupil. A little distance out beyond the eyepiece of a telescope or microscope will be found a round image, which can be caught upon a piece of paper. In a well designed instrument this is a real image of the opening of the objective lens, which is formed by the eyepiece (Fig. 33–8), and it is known as the *exit pupil,* because all the rays that enter the objective go out through it. In order to see the entire field of view of the telescope at once the eye should be placed at the exit pupil. If the exit pupil is too close to permit this (as may happen when the observer is wearing spec-

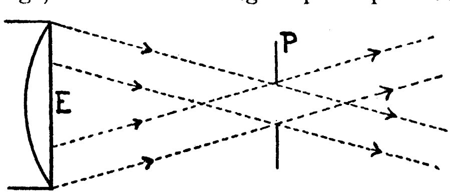


FIG. 33–8
The exit pupil

tacles), only a portion of the field can be seen at one time. If the exit pupil is some distance in front of the eye, the resulting narrowness of the field may become very objectionable.

The magnifying power of a telescope. The magnifying power of a telescope is easily found. The instrument produces first a small real image I_1 formed by the objective and then an enlarged, virtual image I_2 of this, formed by the eyepiece, as indicated in Fig. 33–9. The geometrical relations between the images are here shown, ***not*** the rays of light forming them. What we call the magnifying power is the ratio of the angle A subtended by the final image at the observer's eye, to the angle a subtended by the object itself. The angle subtended by the object is its actual size divided by its distance, and Fig. 33–9 shows that this is the same as the angle subtended by the first image at the objective, or if I_1 is the

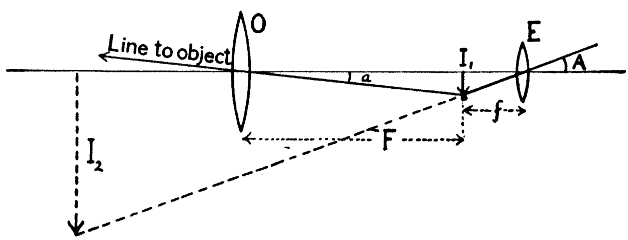


FIG. 33–9

size of the first image and F the focal length of the objective, this angle is approximately I_1/F in circular measure (radians). The angle subtended by the final image is equal to I_2 divided by its distance from the eyepiece, which is equal (by similar triangles) to I_1/f, where f is the focal length of the eyepiece. Hence the ratio of these two angles, which is the magnifying power, is

$$\frac{I_1}{f} \div \frac{I_1}{F} = \frac{F}{f}.$$

This is an interesting quantity. To enable one to see distant objects with any desired degree of enlargement one has apparently only to increase F and diminish f. Practically, F in the Yerkes telescope has attained the enormous length of 1900 cm. and f may easily be made half a centimeter or less. Hence a power of 4000 can be produced.

In using a telescope, however, one readily discovers that the air is a medium through which it is usually not possible to see clearly

at a distance. Variations in temperature and moisture in the air cause the rays of light passing through it to bend irregularly. Everyone knows how the air seems to quiver over a hot level surface out of doors, so that no clear vision through it is possible. A good telescope cannot be kept in a heated dome in winter because the wavering of the air through the opening would ruin the "seeing." The flickering of stars when low in the heavens is due to the same cause and can be seen with the naked eye. With a good telescope conditions of good seeing are so rare that a power greater than 2000 can rarely be used in astronomical work; and a person stalking big game may find that there is no advantage in going beyond a power of 50.

The phenomena of diffraction (p. 551) also set a limit to the possible magnifying power of a telescope, a limit inherent in the wavelike nature of light itself.

Eyepieces. The eyepiece of any optical instrument is not usually so simple as has so far been indicated, but consists of at least two lenses. The advantages secured are a wider field of view and a greater freedom from optical defects. Two types are in common use. The *Ramsden eyepiece* (Fig. 33–10) consists of two plano-convex lenses, the curved sides of which face each other. The lenses usually have nearly equal focal lengths and are separated by a distance a little less than the focal length of either. A real image formed by the objective lens of the instrument at I, just in front of the front lens F, will then be in comfortable focus to an eye placed at E; that is, parallel rays will enter the eye.

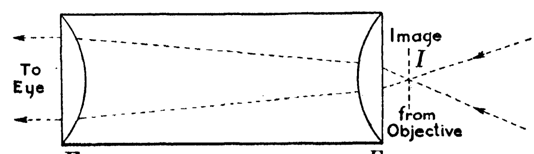


FIG. 33–10
The Ramsden eyepiece

Cross-hairs may easily be used with this form of eyepiece as they can be mounted outside it in the plane of the image I.

The *Huygens eyepiece* consists of two unequal plano-convex lenses with their plane sides toward the eye. The lens nearer the eye has half the focal length of the other. The real image formed inside the instrument of which this is the eyepiece is now between these two lenses, as shown (Fig. 33–11), and cross-hairs cannot quite so conveniently be mounted there. The Huygens eyepiece is the type most commonly used in telescopes and microscopes.

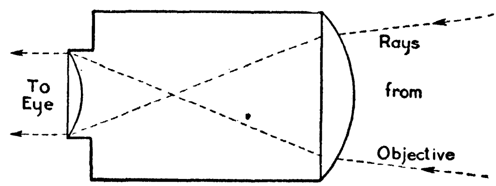


FIG. 33–11
The Huygens eyepiece

Both types of eyepieces are partially free from color defects, but not fully

so. Better eyepieces have been designed, but are not very commonly used in scientific instruments. Figure 33–12 shows a simple triplet magnifier with good color correction (though not a very wide field of view) which is being used very satisfactorily in certain new spectroscopes.

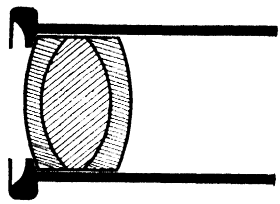


FIG. 33–12

The opera glass, or Galilean telescope. In the opera glass (or Galilean telescope) the rays from the objective are intercepted before they come to their focus by a diverging lens, which causes them to emerge parallel and thus enter the eye (Fig. 33–13). Since the rays have not crossed, the image is erect. The telescope is short and compact and may give sharp images, though over a narrow angle of view. Galileo made one with a magnifying power of 33, but they are usually restricted to 6 or less. The exit pupil (p. 514) is the image of the objective formed by the eyepiece. As this is virtual, it lies between the two lenses where the eye cannot

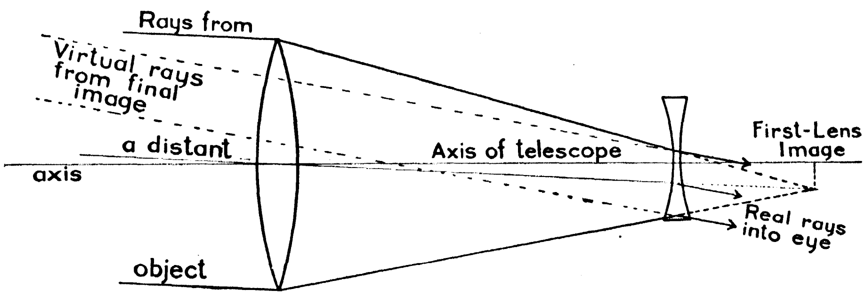


FIG. 33–13
The opera glass

be placed. The eye sees the field of view through this small hole, as it were, and thus the field is very seriously restricted unless the objective is large and the magnifying power very low.

The microscope. A *simple microscope* is just a single lens of short focus used as a magnifier. The earliest ones of high power were probably formed from small drops of molten glass which happened to solidify into a good shape. Some such lenses, improved by polishing, were probably the ones used by Leeuwenhoek [1]

[1] A. van Leeuwenhoek, 1632–1723, a Dutch microscopist, who first accurately described the red corpuscles of the blood and worked out the life histories of many insect parasites and other minute forms of life. He was supposed to have a secret method of grinding his lenses which he refused to divulge.

with great success in his biological discoveries, though he also combined short-focus lenses in groups and probably knew the compound microscope which had been invented many years before.

The ***compound microscope*** is now used whenever high powers are required. It consists of two short-focus lenses, the objective

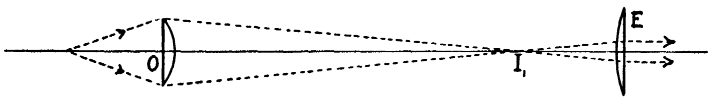


FIG. 33–14
The course of the central rays in a microscope

lens and the eyepiece. The small object to be examined is placed near the objective O (Fig. 33–14), but still beyond its focus, so that this lens forms an enlarged real image I_1 inside the tube of the instrument which is then further magnified by the eyepiece, acting as a simple magnifier. Figure 33–15 shows the actual course

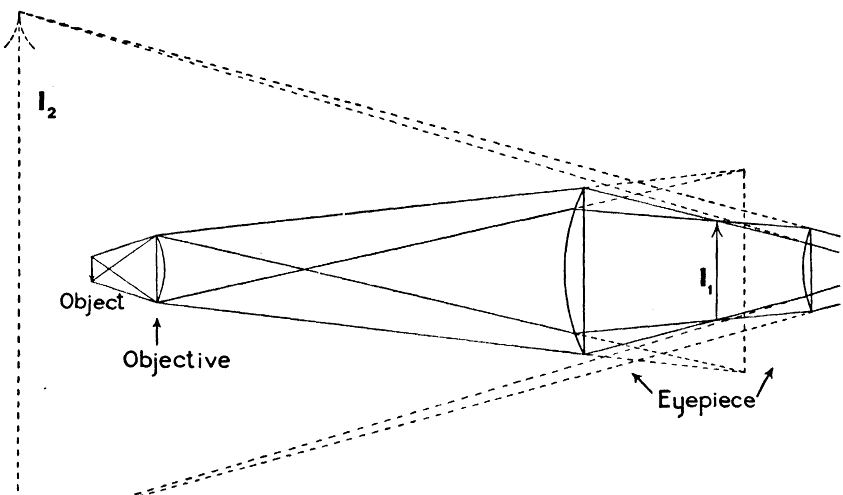


FIG. 33–15
The compound microscope

of the rays starting from two points of the object, forming a real image in the tube (inside the Huygens eyepiece in this case) and a final virtual image beyond the objective on the side away from the eye. It shows the increase in the angle subtended by the final image at the eye compared with that of the object alone, upon which the magnifying power depends. The distance of the final

image is a matter in the hands of the observer. For prolonged use it is comfortable to place it several feet away, except in the case where the observer is making drawings of what he sees with one eye at the microscope, the other eye at the same time showing him his sheet of paper. This somewhat difficult operation is often carried out by biologists, and it is necessary in such cases that the microscope should be so focused as to place the final image at the same distance as the paper. For ordinary use it is often stated that the final image should be placed at the distance of "most distinct vision", i.e., about 25 centimeters or 10 inches. As the size of the final image is proportional to its distance from the eyepiece (p. 509), it looks no larger when near than when far away, and most observers gain greatly in comfort by placing this image farther away than 10 inches.

The early compound microscopes suffered greatly from the defects of spherical and chromatic aberration. In good instruments at the present time the objective is a complex group of lenses whose equivalent focal length may be very short. Thus it becomes possible to place the object very close to the first lens surface and still have it beyond the focal distance, as it must be if it is to form a real image inside the tube. The objectives of the highest powers are the so-called "oil-immersion" lenses with which a drop of oil is used to fill the space between the lens and the thin sheet of glass covering the object. As the oil has the same index of refraction as glass, two refractions are thereby avoided, and a better correction of optical defects is made possible. Water-immersion lenses are also sometimes used.

The magnifying power of a microscope. The real image formed by the objective lens is larger than the object in the ratio of their distances from this lens. Thus if q is the distance from the objective (assumed to be a thin lens) to the image, and p that to the object, the enlargement produced in the first image is the factor q/p. But the first image is further enlarged by the eyepiece, acting in the manner of a simple magnifier, in the ratio $25/f$ (p. 511) where f is its focal length in centimeters; (or $10/f$, if in inches). Thus the whole magnifying power of the instrument is $q/p \times 25/f$ or $q/p \times 10/f$, depending on the units used. To find the numerical value of this quantity in any real case, one needs to measure the distance q, which is easily done, approximately, as it is large. The equivalent focal length of the objective is usually

marked upon it; from these two items the image equation (p. 496) will yield the value of p. The eyepiece often has its magnifying power marked upon it, in which case the magnifying power of the whole instrument is found by multiplying this number by the value of q/p. Some eyepieces, however, are marked with their focal lengths; if so, the full formula above must be used.

Magnifying powers of over 1000 are possible, but there are difficulties in illuminating the object sufficiently to be able to see it when the image becomes so large, as the light originally falling upon the object becomes spread out over a very large area. The phenomena of diffraction (p. 551) enter here likewise, introducing slight defects in the image formed by the objective, no matter how good that is, so that it cannot be magnified indefinitely to advantage.

The smallest visible particles. There is an interesting physical limit to the smallness of particles which are visible, due to the fact that we see by means of waves of light. If the waves in passing by a particle are stopped or otherwise affected by it, the eye, aided by a powerful microscope, may see the particle. But if the particle is small compared with the wave-length, the wave will be unaffected by its presence, just as a water wave will flow past a small stick driven into the bottom in a shallow pond without suffering any alteration of form due to the presence of the stick projecting through the surface. If our information as to the existence of the stick were dependent upon the change it made in the water wave, we should then be unable to detect its presence. In the same way a particle a tenth of a light wave-length in size (i.e., about 0.000005 centimeter or 0.000002 inch) cannot be seen in the ordinary way by any form of microscope, however powerful.

The ultra-violet microscope and the ultra-microscope. If we can use shorter wave-lengths, such as occur in ultra-violet light (p. 537), with which to make our observations, we can observe smaller objects. Such light is invisible, and does not pass through glass, but microscopes have been made with quartz lenses and fluorescent screens (p. 603) on which to receive the images, and by such means a factor of about two has been gained in the smallness of objects that can be observed. This seems a very little result for so much trouble, but if the germ of a virulent disease happens to have such a size that it can just be studied in this way, and not at all otherwise, the result may be very important.

The **ultra-microscope** acts in quite a different way. Referring again to the stick projecting through the surface of the pond, we shall find on close examination that very small ripples are generated at the stick even though the

main waves go by unchanged in form. These ripples die out soon and look very unimportant. In somewhat the same way a particle which is small compared with the wave-length of light scatters a minute amount of light which can best be observed at right angles to the light beam. The particle then looks like a little star emitting (apparently) a feeble light of its own against a dark background. A smaller particle will show as a fainter star *of the same size.* The shape of a small particle is not indicated in any way. A large evaporating drop will show its shape perfectly until it shrinks to about a wave-length in diameter, after which it will become fainter, but remain the same apparent size. The reason for this is that the image of a small object is after all only a diffraction pattern, characteristic of the aperture of the microscope and the wave-length of the light, but not of the object (p. 551).

The projector. The projector, or projection lantern, as used for throwing enlarged images of slides or films on a screen, consists of two parts, the illuminating device and the lens which forms the image. In Fig. 33–16, S is a strong, concentrated source of light (arc, or projection lamp), C is a pair of "condenser" lenses which

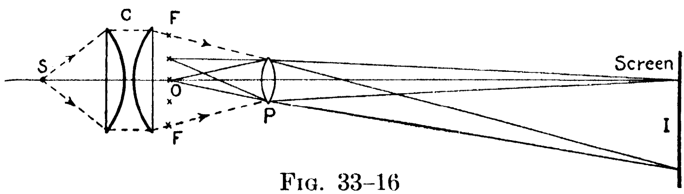


FIG. 33–16
The optical system of the projector

gather up a large amount of light and send it through FF (the film or slide) in such a direction that it all reaches the projection lens, P. The point O on the film is one of the objects whose image I is cast by the projection lens on the screen in the familiar way. By choosing a short-focus projection lens a large image may be thrown on the screen. The dotted lines in the figure indicate the outer boundaries of the illuminating beam; the continuous lines show the image-forming rays, which come, of course, from the source, but are modified by the object so that we may regard them as starting out independently there. If the illuminating rays do not all fall on the projection lens the whole picture will not be uniformly lighted.

PROBLEMS

1. How is a simple magnifier to be adjusted for best use? Where is the final image, and of what character is it? Find the distance of the final image if the magnifier has a focal length of 2 cm., and the object is placed (*a*) 1.9 cm. from it, and (*b*) 2.1 cm. from it.

2. Design an eye suitable for a predatory animal, so that it can see equally well above and under water, without any change of adjustment.

3. A Galilean telescope consists of a converging lens of 12 cm. focal length, and a diverging lens of 4 cm. focal length. The two lenses are 4.1 cm. apart. Where is the final image of a distant object as seen in this instrument?

4. An achromatic lens consists of a combination of two lenses, of + 6 in. and − 10 in. focal length respectively, placed 0.5 in. apart. Find the position and character of the image, formed by this combination, of an object 24 in. away from the converging lens.

5. A pocket magnifier consists of two converging lenses, of 1 and 2 cm. focal length respectively, placed 0.5 cm. apart. Where is the final image of an object formed by the magnifier when that object is placed at a distance of 1.5 cm. from the 1 cm. lens? Of what sort is this image?

6. An achromatic lens is made up of a positive lens of 20 cm. focal length, and a negative lens of 30 cm. focal length placed 2 cm. behind the positive lens. Find the position of the final image, formed by this combination, of an object 100 cm. in front of the positive lens.

7. Find the position of the final image of a distant object formed by a pair of lenses consisting of a converging lens of 10 cm. focal length, and, at 1 cm. distance beyond, a diverging lens of 20 cm. focal length.

8. Find the position of the final image of an object 2 m. away from the objective lens of a telescope (20 cm. focus) which is in its turn 21 cm. away from a diverging eyepiece of − 5 cm. focal length.

9. An astronomical telescope has an objective lens of 30 cm. focus and an eyepiece of 3 cm. focus. These lenses are 34 cm. apart. An eye looking through this telescope, as so adjusted, at objects both far and near finds that those at a certain distance are clear and distinct. What is this distance?

10. A prism field glass has an objective of 24 cm. focal length and an eyepiece of 3 cm. focal length. If the sun shines into the objective, a distinct, enlarged image is formed 50 cm. beyond the eyepiece. What is the distance between the lenses?

11. A microscope has an objective of 2 cm. focal length and an eyepiece of 3 cm. focal length. The object is 2.2 cm. from the objective. Find the proper distance between the lenses for comfortable use, and the magnifying power with this adjustment.

12. An astronomical telescope consists of an objective lens of 50 cm. focal length and an eyepiece of 2 cm. focal length. If this telescope is pointed toward an object 2 m. away, how far apart must the two lenses be in order that the object will be seen clearly through the instrument?

13. An astronomical telescope whose objective lens has a focal length of 20 in. is pointed toward a lamp 20 ft. away. The eyepiece has a focal length of 1 in. and is placed 21 in. from the objective. Find the position of the final image as formed by the telescope in this state of adjustment. Will it be

satisfactorily clear to the eye, if one looks into the eyepiece in the usual manner?

14. An opera glass has an objective lens of 4 in. focal length, and an eyepiece of 1 in. focal length. It is desired to point this toward the sun and cause it to project a real image of the sun on a card beyond the eyepiece. The eyepiece is 3.03 in. from the objective. Where is the final image?

15. A compound microscope has an objective of 1 cm. focal length, and the object is placed at a distance of 1.1 cm. from it. The eyepiece has a focal length of 3 cm. and the two lenses are 15 cm. apart. Find the nature and position of the final image. Is the instrument in proper adjustment for ordinary use under these conditions?

16. Compare the brightness of the images given by two field glasses, if one has a power of 8 and an objective lens 28 mm. in diameter, while the other has a power of 6 and a 24 mm. objective.

17. By using the image equation and the facts stated on page 514, prove that the magnifying power of an astronomical type of telescope (or a prism field glass) is equal to the ratio of the diameter of the objective to that of the exit pupil.

CHAPTER 34

DISPERSION AND SPECTRA

Dispersion by a prism. It is interesting to repeat an experiment done by Newton, in which he placed a prism in a narrow beam of sunlight coming through a hole in a shutter, and bent it so that it fell upon a wall as a strip of colors which he called a ***spectrum*** (Fig. 34–1). The violet end of this strip was bent through the largest angle by the prism; the red the least. The colors of the spectrum are often said to be red, orange, yellow, green, blue, indigo, and violet. When one makes a fairly large ***pure*** spectrum (p. 528), i.e., one in which the colors do not overlap, it is seen that these colors shade imperceptibly into one another, so that seven colors are insufficient to describe what one sees. Over a hundred (shades, if not colors) can actually be distinguished by a normal eye in a properly arranged experiment. There is thus nothing sacred about the seven spectrum colors; indeed, most observers omit the indigo from the list, feeling that the distinction between it and the blue on one side and the violet on the other cannot be made, at least by ordinary eyes.

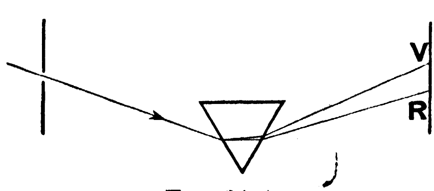


FIG. 34–1

The separation of the colors in such an experiment is an example of what is called ***dispersion.*** The order of the colors proves to be

the same as the order of the lengths of the light waves (p. 555), the longest waves being deep red, the shortest violet. The actual width of each color absolutely and relatively to the other colors depends on the material of the prism. In general those substances which bend the beam through the greatest angle, i.e., have the greatest index of refraction, also disperse the colors into the widest spectrum. Newton thought there was a fixed proportion between the refraction and the dispersion, the same for all substances. That this is not true is important, as it has led to the possibility of making good achromatic lenses and refined optical instruments generally. Certain sorts of flint glass which have been produced in recent years give a relatively large dispersion for a given angle of refraction when compared with ordinary, or crown glass.

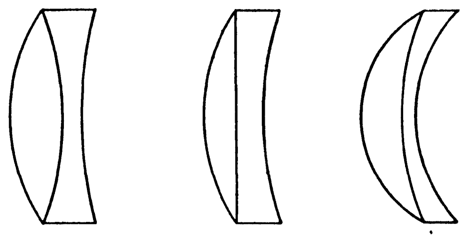


FIG. 34–2
Achromatic lenses of different forms

The achromatic lens. A lens is said to be achromatic ("without color") when it gives images in which all colors come to the same point, without the colored edges which are so commonly seen when uncorrected lenses are used.

If a strongly converging crown glass lens is combined (Fig. 34–2) with a diverging flint glass lens which produces the ***same amount*** of dispersion of the colors as the crown glass lens does but in the ***opposite*** direction, the refraction of the crown glass lens will be only partially annulled, and the combination will act as a weakly convergent lens, bringing (practically) all the colors to a focus at the same place. This action can perhaps be more easily followed by considering an analogous, though less useful, device, the ***achromatic prism.*** *C* in Fig. 34–3 is a crown glass prism upon which a narrow beam of light falls which would emerge in the direction *E* if there were not a flint prism *F* in the way. The beam

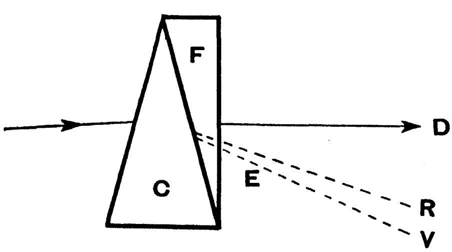


FIG. 34–3
An achromatic prism

E would produce a spectrum of a certain width *RV* on a screen. But on account of the presence of the prism *F*, the beam emerges in the direction *D*, the refraction of the flint prism bending the colors *RV* back to a single line. Since the refraction of the flint prism is not so great as that of the crown, there remains a bending of the beam in the direction demanded by the crown glass, but this refraction is colorless, or achromatic, since *R* and *V* (and all the other colors) have come together at *D*.

In the achromatic lens (Fig. 34–4), the crown glass lens *C* alone would concentrate an incident parallel beam of white light *L* into a cone of violet rays converging at a violet focus *V*, and a similar but longer cone of red rays converging at *R*, together with a multitude of other cones overlapping between, produced by the intermediate colors. The presence of the flint lens, however, slows down the convergence of the violet rays more than that of the red, and what really happens is that all the colors go together to form a white image at *W*. (Note also Example 1 on p. 512 where the focal length of such a combination is calculated.)

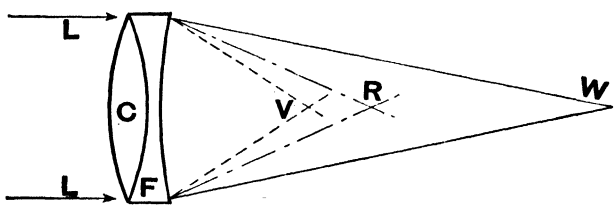


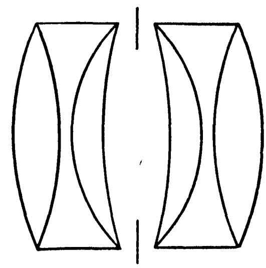
Fig. 34–4
Action of achromatic lens

With two lenses combined, *two* colors may be brought accurately to the same focus. As a rule, however, the other colors do not come precisely to the same place. The dispersions of the two sorts of glass are not usually quite in proportion, the green, for instance, being relatively wider in a spectrum formed by a prism of one sort of glass than in one formed by another. This peculiarity is sometimes given the formal name of the "irrationality of dispersion." The result is that when two colors are chosen to be brought together, say the bright red and blue, the deep red and the violet commonly focus farther out, and the yellow and green slightly closer than the main focal point. In astronomical telescopes a violet halo is often seen about the images of bright stars which is due to this cause. To get rid of this slight defect in a telescope (or a microscope) lens, at least three different lenses must be combined into one, and then three colors can be brought to the same

focus, with the result that no other color can go appreciably astray. Such lenses are said to be "apochromatic."

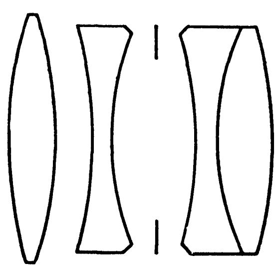
Photographic lenses. So far we have assumed that achromatic lenses are intended for visual purposes. If they are to be used for photography, new conditions arise. The ordinary photographic plate is most sensitive to the violet part of the spectrum and least to the yellow and red. The eye, on the other hand, is most strongly affected by the yellow-green, and very little by the violet. The common photographic film is treated so as to make it sensitive to the yellow, blue, and violet, and it is possible to obtain photographic plates or films which are *panchromatic*, i.e., sensitive to all the colors, "seeing" them much as the eye does. Hence a really good photographic lens should be corrected for all colors, and this is generally done.

Photographic lenses have to meet tests which are unusually rigorous. An ordinary telescope lens is required to form a very sharp image along its own axis only, but a good photographic lens must furnish sharp images when the rays of light go through it at a considerable angle, as they do to form the corners of the average photograph. A common lens suffers from the defect called *astigmatism* (p. 502) when forming images by means of rays passing through it at a considerable angle. To correct a lens for all its defects is difficult; as astigmatism is perhaps the worst, such lenses are usually called anastigmatic (literally "not-not-point-forming"), meaning that they *do* form point images of point sources on any part of the photographic plate which they are intended to "cover," and do this for all colors. Figure 34–5 shows the construction of two celebrated lenses of this nature, one made up of three components, the other of two symmetrical triple lenses.



a

The Goerz "Dagor" lens



b

The Zeiss "Tessar" lens

FIG. 34–5

The best lenses of this sort are also corrected for *distortion* (lack of rectangularity in the image of a perfectly rectangular object) and *curvature of field* (the image of a flat object not lying all in one plane), as well as for the defects already mentioned. So much care must be spent on their construction that they are necessarily rather expensive.

The spectroscope. A vast amount of information has been accumulated from the study of spectra, which is useful in chemical analysis, in the study of the physical state and the motion of the sun and stars, and in physics especially in relation to the problems of atomic and molecular structure. All this has been gathered by

means of ***spectroscopes*** of one form or another. We must consider the essential features of these instruments.

Newton's simple arrangement for producing spectra (Fig. 34–1) has the defect that the round spot of light formed by each color overlaps that of adjacent ones. Figure 34–6 represents a succession of such round images, each of a different wave-length, or shade of color, spread out into such a spectrum as Newton saw.

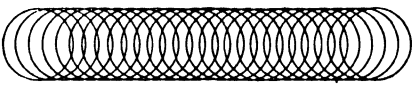


FIG. 34–6

The overlapping which this discloses may be avoided by limiting the beam sideways by means of a slit. There should also be a lens to throw an image of this slit upon a screen, and, of course, a prism to cause the images of the slit which are formed in the different colors to separate out and form a spectrum. In the better instruments two lenses are used, one on each side of the prism, which gives slightly better definition. One of these, C, (Fig. 34–7) is called the "collimating" lens, and is so placed that rays from the slit passing through it emerge parallel. After being bent by the

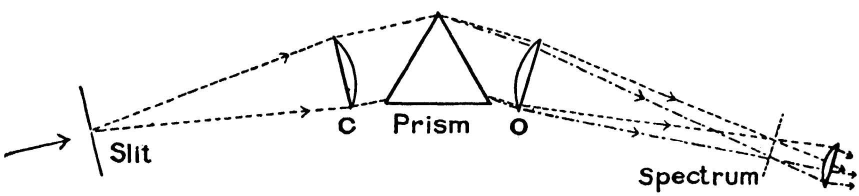


FIG. 34–7

prism the rays of each color are still parallel but there is one set of them for each wave-length that happens to be present, and each set goes in a slightly different direction. Two sets are shown in Fig. 34–7. These rays enter the observing telescope O and are brought to the eye as in any other telescope. This telescope can usually be turned about the prism as center so as to receive in succession all the parts of the spectrum. What is seen, then, is an image of the slit for each wave-length that is present, and these are so narrow that they do not overlap. A spectrum in which only one wave-length reaches any given spot is called a ***pure*** spectrum.

Direct-vision spectroscope. The ordinary spectroscope involves an angle between the incident and emergent rays which is sometimes awkward. A pocket form of spectroscope is in common use

in which this angle is avoided. It contains a multiple prism as in Fig. 34–8 in which three prisms of crown glass are cemented to two of flint. The angles and dispersing powers of the glasses are so chosen that the green rays go straight through, while the red are bent to one side and the violet to the other. In other words the refraction is neutralized while the dispersion is not; exactly opposite to what occurs in the case of the achromatic lens.

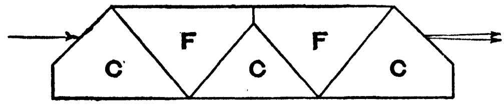


FIG. 34–8

These little instruments are also often made with gratings (p. 552) rather than prisms to form the spectra.

Prism spectrographs. A spectroscope may be altered into a photographic instrument by removing the observing telescope and substituting a camera, preferably with a long focus lens. Such an instrument is called a *spectrograph*. It is often advantageous to photograph a spectrum rather than observe it by eye only. Measurements can be made on it more readily; very faint lines may be obtained by prolonged exposures, and permanent records may be obtained of the spectra of temporary sources (e.g., flashes of lightning) which one cannot have under control, or repeat at will.

Types of spectra. It is fascinating to send light from one source after another into a spectroscope and discover the types of spectra that can readily be produced. The commonest is the one yielded by all hot solids and liquids, which is a ***continuous spectrum,*** a perfectly unbroken strip changing imperceptibly from one color to the next. The common gas-filled tungsten-wire incandescent electric lamp gives this spectrum particularly well. There are no such things as natural gaps between the colors. In terms of the description given above there is an image of the slit for every wavelength in a continuous spectrum, and these fit so perfectly next to one another that there are no gaps anywhere. Figure 34–9*a* shows a photograph of such a spectrum as seen in a prism spectroscope.

While this spectrum is being observed, if one places a piece of cobalt-blue glass in front of the slit of the spectroscope, the spectrum is as shown in Fig. 34–9*b*; the orange, yellow and green are cut out. A piece of "signal" green glass gives Fig. 34–9*c* showing green only; "didymium glass" gives Fig. 34–9*d* showing absorption in several places, densest in the yellow. These spectra are of the class known as ***absorption spectra;*** they are formed from

a continuous spectrum by sending the light through an absorbing medium, which may be almost any colored solid or a solution held in a clear glass cell. The absorption "bands" disclosed by the spectroscope are usually wide when produced by solids or liquids, though in rare cases (e.g., Fig. 34–9*d*) they may be rather narrow, in which case the substance may not appear to have much color. When absorption is produced by gases, it usually takes the form of very narrow bands or lines. Figure 34–13*b* shows a simple case, the absorption spectrum of a sodium flame. These lines may be very numerous; one of the best examples is furnished by the spectrum of the sun Fig. 34–13*a* (p. 540).

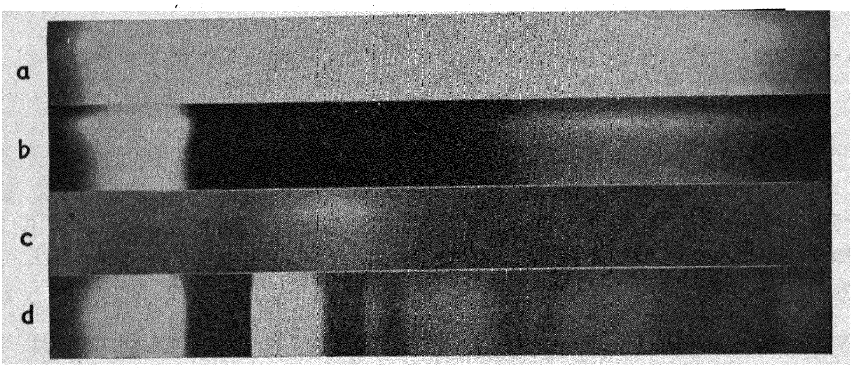


FIG. 34–9

Continuous and absorption spectra, red on the left, violet on the right

A third type of spectrum is seen if we replace the lamp by a Bunsen flame colored by the addition of common salt (sodium chloride). The flame itself should be blue, emitting almost no light of its own. The salt may be introduced by soaking a bit of asbestos in it and then bringing this into contact with the lower part of the flame. We then see in the spectroscope one of the simplest of the ***bright-line spectra,*** consisting of a single image of the slit of a yellow color, due to sodium, or, if the spectroscope is a very good one, and the slit is narrow, we may see that it is really a double image consisting of two close spectrum "lines," one a little stronger than the other. The rest of the spectrum is simply not there; the "background" is black, or only very faintly illuminated by the blue and green rays from the flame itself, which are often invisible (Fig. 34–10*a*; note that the photographs in this figure are reproduced as negatives, so as to

show the lines better; thus the one yellow line on a black background becomes a single black line on a light ground). If a little lithium chloride is then put on the edge of the same bit of asbestos, the spectroscope will show an added red line (Fig. 34–10*b*). A little solution of a thallium salt will contribute a green line (Fig. 34–10*c*); and so on. Any one substance always produces a line (or lines) of the same color; in fact, a line in ***exactly*** the same part of the spectrum, as indicated by accurate wave-length measurements considered below (p. 555).

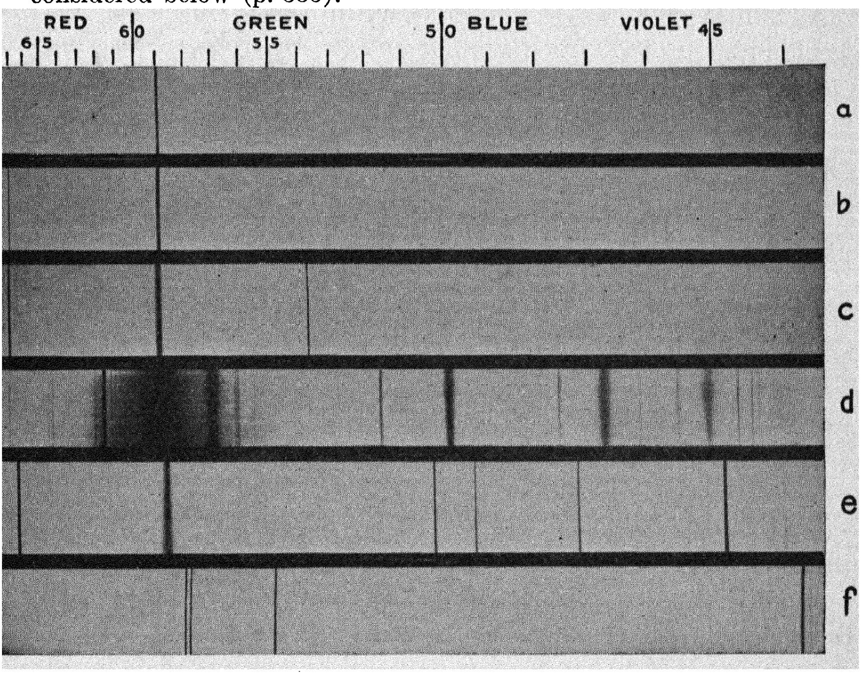


FIG. 34–10

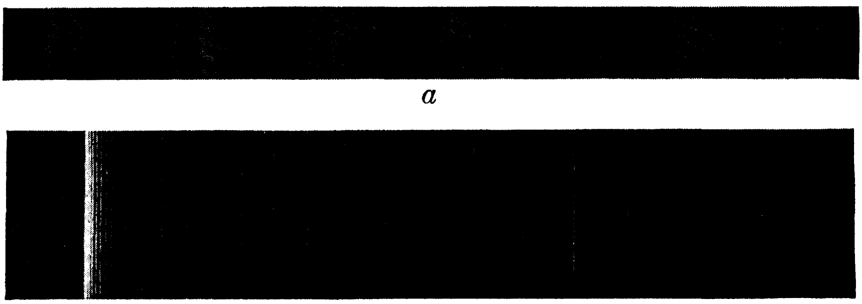
Bright-line spectra, here printed reversed, as photographic negatives

More interesting bright-line spectra may be produced by the ***electric arc.*** Ordinary arc carbons may be soaked in a strong solution of common salt. The arc is then formed between them, and an image of this arc is thrown on the slit of the spectroscope in such a manner that the light entering the instrument comes from the vapor between the carbons, rather than from the hot carbons themselves. There will be seen in the spectrum a rich and beautiful array of bright lines of various intensities (Fig. 34–10*d*), most of which are produced by the sodium in the salt. The yellow lines which were present in the flame spectrum (*a*) are very strong, and

in the flame (Fig. 34–10*a*) is very striking. We shall return to these variations again, as they yield much information about the constitution of atoms and molecules (p. 548).

Many attempts have been made to put spectrum analysis on a quantitative basis, so as to be able to say from the spectrum what percentage of a given substance was present in the source. While this can be done in many cases with moderate precision, each mixture of substances furnishes a new problem requiring separate study.

Band spectra. There is another sort of bright-line spectrum which is frequently observed, and which is called a *band spectrum.* A typical one is shown in Fig. 34–11*a*. When on a small scale they remind one of a fluted



a

b

Fig. 34–11

Band spectra; (*a*) spectrum of boron oxide; (*b*) one band in the spectrum of the ionized nitrogen molecule, here greatly enlarged

column in appearance, and the name "fluted spectra" used to be applied to them. When they are made on a large scale, they are found to consist of multitudes of lines, arranged in a very orderly and characteristic fashion (Fig. 34–11*b*). Band spectra have recently been shown to be formed by the vibration of molecules, whereas ordinary line spectra are due to atoms. The lines in the bands arise from to-and-fro vibrations of the atoms in the molecules, or from their spinning around one another, or from the activities of electrons within the atoms, or from all three of these actions together. The study of the arrangement of the lines in bands has led to a great increase in our knowledge of the nature of molecules, and of the forces holding them together. In many cases these molecules prove to be incomplete; or of a sort which could exist only momentarily, due to some temporary peculiarity of one or more of the atoms in them.[1]

Bands may be observed either in bright-line or in absorption spectra. They

[1] For instance, the molecules He_2, Hg_2, NaK, CaH, HO, CH, etc. are not met with in chemistry, but have been shown to exist, at least briefly, in light sources.

are commoner in absorption, because the substance producing an absorption spectrum is usually comparatively cool, while in sources of bright-line spectra the particles are so agitated that a great many molecules would be broken up into their constituent atoms and could no longer hold together.

The invisible part of the spectrum beyond the red. It is not difficult to show that the spectrum formed from a bright source by a prism spectroscope does not end at the edge of the red as it appears to when viewed in the ordinary way. First, this limit can be pushed out considerably by using a strongly concentrated source such as an arc and filtering out all the rays that affect the eye most strongly by means of a piece of cobalt-blue glass placed in front of the slit. This allows the deep red to pass through it (as well as the blue) and stops the bright red, orange, yellow, and green, whose presence ordinarily makes the eye less sensitive. In their absence the spectrum can be seen to stretch out considerably farther than would otherwise be supposed. But if an instrument such as the thermopile (p. 368), which is sensitive to the presence of minute amounts of heat radiation, is placed in the spectrum, it is found that the heat received by the instrument increases rapidly as it passes from the yellow through the red into the dark region beyond, usually called the ***ultra-red.*** It usually rises to a maximum there and then falls again as the instrument is pushed still farther out. Figure 34–12 shows the distribution of energy among the different wave-lengths as determined by careful measurements of this sort, using as source a hot, solid "black" body at a series of different temperatures. These spectra are all of the type known as continuous spectra. It is easily seen that the energy is not uniformly distributed among all wave-lengths, but rises to a well-defined maximum. If the maximum energy in a curve obtained from a source at an absolute temperature T occurs at a certain wave-length λ_m, these measurements have shown that the product $\lambda_m T$ is a constant, which has the value 2885 if λ is measured in microns (1 micron = 1 μ = .001 mm.). Thus at a temperature of 373° K. the maximum energy is given out at a wave-length of 7.7 μ in the deep ultra-red; at 2800° K. at 1μ in the near ultra-red and at 6000° K. at 0.48μ in the blue-green. This law has been derived from theory as well as experiment, and we need not hesitate to trust it to hold over a wide range of temperatures. In fact it furnishes an excellent basis for a method of measurement of very high temperatures, i.e., a ***radiation pyrometer.*** While it is a some-

what complicated experiment to find in which part of the spectrum the greatest energy (i.e., heat radiation) lies, still, when this has been determined, the calculation of the temperature of the source is simple and trustworthy, and it is a relief not to be bothered by the limits such as are set in ordinary high-temperature thermometry by the melting of thermometer bulbs, etc. By this method the average temperature of the sun's surface has been estimated at 6000° K., after making due allowance for the effects of atmos-

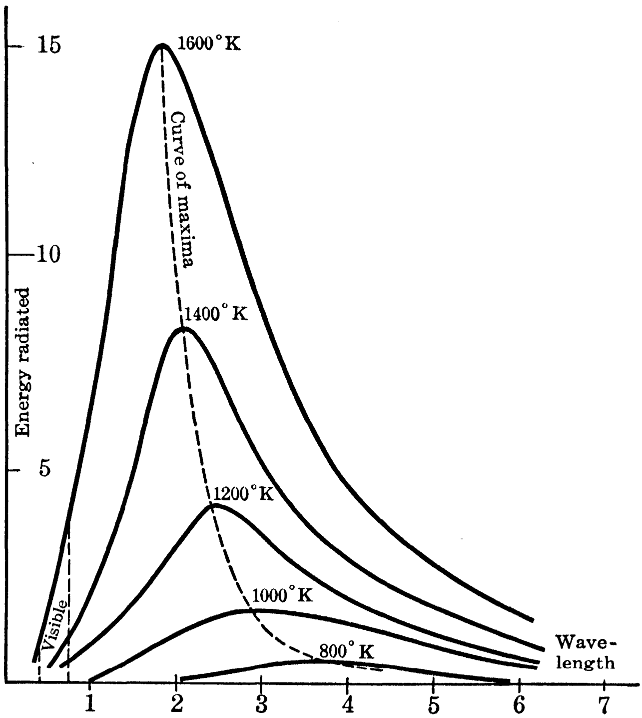


FIG. 34–12
Energy curves

pheric absorption on the energy as received by us. The λ_m corresponding to this figure is in the blue, indicating that the sun would be a bluish-white if seen by an observer outside of our atmosphere. Temperatures up to 20,000° K. have been determined for certain bright stars by the same method.

Ultra-red absorption spectra. Beside the continuous spectra mentioned above, ultra-red absorption spectra have also been observed. Many substances absorb the long, "heat" waves; the air itself does this, mainly because of the presence in it of water

vapor and carbon dioxide, both of which absorb powerfully certain limited parts of the ultra-red. Glass transmits only the shorter, or near, ultra-red rays. A spectroscope with glass parts is unsuitable for the study of the ultra-red region; rock-salt crystals must be shaped into lenses and prisms for this purpose, as this material is transparent over a wide range. A sheet of hard rubber is transparent to the longer ultra-red rays, though quite opaque to ordinary light; hence lenses of hard rubber can be made to focus them. It is not difficult to show that rays emitted by a red-hot body can go through a sheet of hard rubber and warm a thermopile beyond, but these same rays can be stopped by placing a piece of glass in the way. On the other hand, some rays get through glass which are stopped by the hard rubber. It is a question of their wave-lengths. Glass absorbs waves whose lengths are greater than about 1.5μ (0.00015 cm.) and hard rubber transmits some of those waves which are too long to go through glass.

There are special eye-protecting goggles made for men to wear while working near masses of molten iron, etc. These do not allow the ultra-red rays to reach the eye. Usually these glasses are not absorbent, but are coated with a thin film of metal which, though almost transparent for ordinary light, ***reflects*** the ultra-red powerfully.

The greenhouse problem. It is interesting to see how these ideas are illustrated by two very practical cases, one of which is furnished by the greenhouse. It is a well-known fact that a greenhouse acts as a trap for the sun's heat, and will reach a higher temperature than the air outside when the sun shines. The reason is that the greater part of the sun's energy is borne to us by short waves which pass freely through the air and the glass roof of the greenhouse. This heat falls upon the objects within and warms them. As a result these objects begin to radiate waves which must be very long, since the temperature is not high; but glass is opaque to these waves, and thus absorbs or reflects them, and retains most of the heat within the enclosure. It is easy to cook an egg inside a glass-topped box which is provided with heat insulation on its sides and bottom, and is exposed to the sun's rays.

Precisely the same action accounts for the warming of the surface of the earth by sunshine. Since the sun's heat enters freely, it warms the ground, but much of the returning long-wave radiation which the ground sends out in turn is absorbed by the water

vapor and carbon dioxide in the lower layers of the atmosphere, and hence is trapped. Thus it comes about that the lower parts of the atmosphere are much warmer than the upper. The temperature ten miles up above the tropics is uniformly lower than $-50°$ C. (62° below zero Fahrenheit), as shown in Fig. 2–5 (p. 20).

Ultra-red emission spectra. Many substances emit bright-line spectra when placed in a Bunsen flame. The commonest example is furnished by sodium from common salt. Potassium is a very similar element, but it is heavier. If one puts a bit of asbestos moistened with a potassium chloride solution into the edge of a Bunsen flame, a very deep red "line" (really a pair of lines) is emitted by the flame, and may be seen under favorable circumstances. The still heavier elements of the same family, rubidium and caesium, emit similar lines in the ultra-red, which cannot be seen, but can be detected by means of a thermopile, or may even be photographed if special ultra-red-sensitive plates are used. These are examples of ultra-red "bright-line" or emission spectra. The arc spectra of the elements and compounds contain many important lines and bands in the ultra-red, the discovery of which in recent years has added much to the understanding of spectra.

Ultra-violet spectra. Beyond the limit of the visible spectrum at the violet end there lies another part, called the ultra-violet, in which the wave-lengths are shorter than those of visible light. Here continuous spectra from extremely hot solids may be found, though they are usually feeble; absorption spectra are numerous, and emission spectra, both of lines and of bands (atoms and molecules), are of great importance. The ***range*** of the ultra-violet begins at a wave-length of about 0.00004 centimeters (p. 555) and extends to a limit set by the observing instruments, by the air, or by the source. A glass instrument transmits very little of the ultra-violet, as ordinary glass is opaque to it. If quartz prisms and lenses are used, waves whose length is 0.00002 centimeter can readily be observed by using a photographic plate as the receiving surface. Beyond this limit the quartz and the air itself begin to absorb, but by exhausting the air from the instrument, substituting fluorite parts for quartz, and using special photographic plates, Lyman [1] succeeded, in 1906, in photographing waves of almost one-fourth the length of violet light. By the use of reflecting gratings (p. 554) in a vacuum, he later extended these observations another octave, to borrow a musical term, and Millikan later

[1] T. Lyman, director of the Jefferson Physical Laboratory at Harvard University; author of "The Spectroscopy of the Extreme Ultra-violet," 2nd ed. 1928, (Longmans, Green & Co.) in which this field of work is summarized.

reached waves one-thirteenth as long as the violet, or nearly five octaves beyond the visible spectrum.

These short waves produce and can be detected by various phenomena, such as fluorescence (p. 603), and the photoelectric effect (p. 596). Their ***biological effects*** deserve especial mention. The sun's spectrum is cut off rather sharply by absorption in the upper air at about 0.0000295 cm. wave-length. The rays whose wave-lengths lie between this value and 0.000032 cm. are sometimes called the "vital rays," and seem to be the ones that have most to do with the production of sunburn. They are cut off by ordinary glass, but special kinds of glass are now made which when used in window panes enable these rays to get into our houses. Moderate exposure to them appears to be distinctly beneficial to one's general health. The same rays can be produced by artificial sources, of which the quartz mercury lamp is the commonest. Many of these sources emit with considerable intensity still shorter waves (0.000025 cm. and less) which are "abiotic," that is, they burn violently (especially one's eyes) and kill bacteria and other simple living forms. On this account they are used to sterilize milk and water. Men who have to work with electric arcs, especially in welding iron, must be protected from them by wearing glasses over their eyes.

The nature of line spectra. Each line in a bright-line spectrum has a definite wave-length which can be measured with great accuracy by methods soon to be considered in detail (p. 553). Each wave-length has a definite frequency of vibration. One can make a crude but interesting analogy between a line spectrum and a complex musical sound such as one might make by striking several piano keys at once. Assuming that each piano string gives one note only, we hear a complex sound containing just as many frequencies as there are keys pressed. Moreover, the positions of the keys along the keyboard give some indications of the frequencies, or wave-lengths, because they are arranged in an orderly fashion. A drawing of the keys which are pressed could be called a sound spectrum of the complex note given by the instrument.

Similarly, when the different light vibrations emitted by a flame or arc are sorted out by means of a prism in the order of their frequencies, we find the lines arranged according to a definite pattern for each chemical element.

This analogy must not be pushed too far. At first it was trusted quite hopefully, and the frequencies of the light waves emitted by the atoms were regarded as due to actual vibrations of parts of the atoms themselves; i.e., either fundamental or harmonic frequencies as in sound. It was soon seen that this could not be true, when the actual values of the frequencies were obtained, as the simple numerical connections among the frequencies in the case of musical harmonics had no counterpart in atomic vibrations.[1] Also, several sorts of atoms emit thousands of different frequencies. In a sort of humorous despair it was said that such an atom must be more complicated than a grand piano, which with its 88 keys could scarcely send out a thousand notes even with a most generous allowance for harmonics. Thus the atom stoutly resisted all attempts that were made to explain its optical behavior in simple mechanical terms.

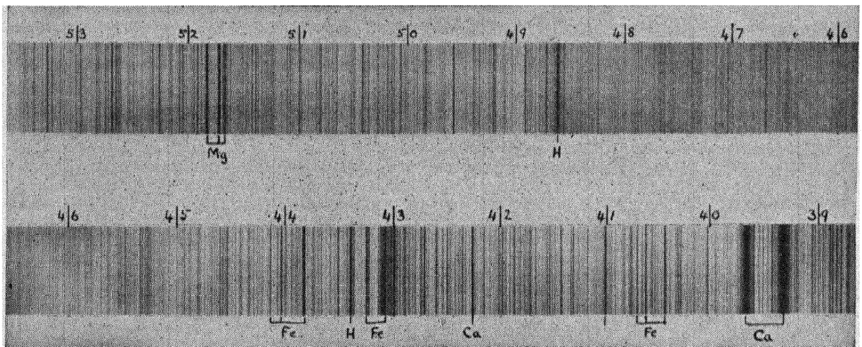
The spectra of the sun and stars. When sunlight is examined in a good spectroscope, it is found to consist of a continuous spectrum crossed with a multitude of dark lines (Fig. 34–13*a*). It is thus the exact reverse of a bright-line spectrum. The meaning of these dark lines is interesting. The body of the sun presumably emits a continuous spectrum, but this light has to pass through the outer atmosphere of the sun and then that of the earth before we can observe it. Each atmosphere removes certain wave-lengths by absorption. In the sun, the hot metallic vapors and gases are in a condition to emit certain frequencies. If waves of these same frequencies, present in the continuous spectrum, fall on these atoms on the way out to us, they are partially absorbed and subsequently re-emitted in all directions. In our direction less energy goes forward on account of this absorption, so that a relatively dark line appears in each place in the spectrum where this happens. Figure 34–13*b* shows the absorption produced in this way in the yellow part of the spectrum, when light from a very hot solid is sent through sodium vapor in a Bunsen flame. It also shows for comparison that part of the solar spectrum which contains this pair of sodium lines. These spectra were taken on a much larger scale than those in Fig. 34–10 or in Fig. 34–13*a*.

This absorption and re-emission reminds one of an experiment on resonance in sound, in which (p. 273) one tuning fork absorbs

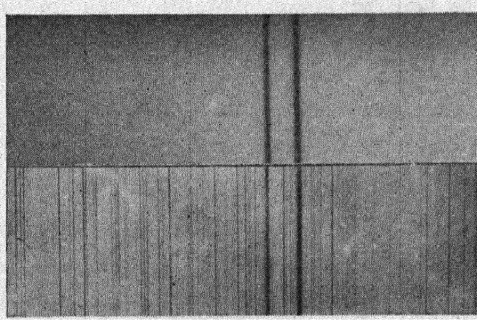
[1] Harmonic relations do occur, however, in the vibrations of certain molecules.

the sound coming from another exactly tuned to it. ***Only those bodies can absorb which are in a condition to emit the same frequency.***

Each dark line in the spectrum of the sun comes from a definite chemical element or compound which is present in one of the atmospheres through which the light has to pass. There are various ways of determining which lines are produced by the earth's



(*a*) The solar spectrum in the green, blue and violet;



(*b*) The same in a small part of the yellow, compared with the absorption of a sodium flame

FIG. 34–13

atmosphere. These lines prove to be due to molecules of oxygen, water vapor, etc. The dark lines produced in the sun's atmosphere itself enable us to perform a chemical analysis on the sun, since ***the wave-lengths emitted by an atom are the same wherever that atom is*** (neglecting very minute changes due to differences in pressure, gravitation, relative velocity, etc.). In this way the presence of fifty-four elements out of our known total of ninety has already been positively established in the sun, and at least four

more are probably there. Over 22,000 lines have been measured in the sun's spectrum, and the origin of about half of these has been discovered.

The most prominent elements there are among the commonest here. The lines of the sun's spectrum were first studied by Fraunhofer[1] in 1815, after whom they are named. He assigned letters to the heavier ones. Thus *A* and *B* are close groups due to oxygen in the earth's atmosphere, *C* and *F* are the red and blue-green lines of hydrogen which constitute the first two of the Balmer series (p. 542), the *D* line is the sodium yellow, (identified as such by Fraunhofer himself) *E* and *G* are mixed groups largely due to iron, *H* and *K* are very heavy lines in the violet due to ionized calcium, *b* is a group in the green composed mainly of three magnesium lines, etc. The wave-lengths of these lines are given in Table XXX (p. 555).

The light from stars, however distant, makes chemical analyses of these bodies almost equally easy. The stars are often so hot, however, that the atoms of ordinary elements are broken up (ionized) by the violent thermal agitation, and the spectra of these ionized elements (p. 548) lie so far in the ultra-violet that they are absorbed by the earth's atmosphere; thus we can find no trace of the presence of these elements in these stars. Such stars appear to consist solely of hydrogen, helium, and a few other elements. It is likely that if they were to cool down, their spectra would show the same familiar elements that we find in the earth and the sun.[2]

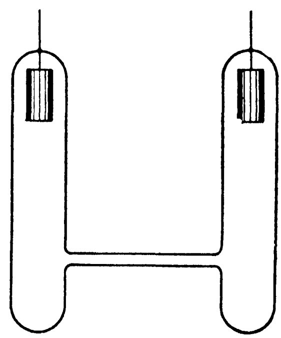


FIG. 34–14

Series of lines in spectra.[3] The spectra of some of the lighter and presumably simpler atoms show lines arranged in a manner which seems to follow a simple law. Hydrogen, for instance, when an electrical discharge is sent through it at a pressure of a few millimeters of mercury in a tube such as is shown in Fig. 34–14, can be made to show a spectrum (Fig. 34–15) com-

[1] Joseph von Fraunhofer (1787–1826), German optician and physicist. Besides devising many useful telescopic appliances, he invented the diffraction grating.

[2] The reader interested in this subject (astrophysics) is referred to "Astronomy" Vol. II (Astrophysics and Stellar Astronomy), by Russell, Dugan, and Stewart, 1927 (Ginn & Co.), a lucid, accurate and inspiring book, written by men who have themselves taken part in the discoveries they describe.

[3] The rest of this chapter may be omitted, if it is desired to save time in a short course.

posed of four lines only in the visible part of the spectrum, but many more in the near ultra-violet, the lines of longer wave-length being widely spaced and strong, the lines of short wave-length coming together as they approach a limit, and fading at the same time. Such an arrangement is called a ***spectrum series.*** Another series of similar aspect has been found by Lyman in the extreme ultra-violet spectrum of hydrogen and others occur in the ultra-red. All these lines, due to atomic (rather than molecular) hydrogen, have wave-lengths which are given by a simple formula

$$\frac{1}{\lambda} = R\left(\frac{1}{n^2} - \frac{1}{m^2}\right)$$

where R is a universal series constant known as the Rydberg [1] constant, and n and m are integers. If n has the value unity and m is given the values 2, 3, 4, etc. in succession, the formula gives the wave-lengths of the lines in Lyman's series. If $n = 2$, and $m = 3$,

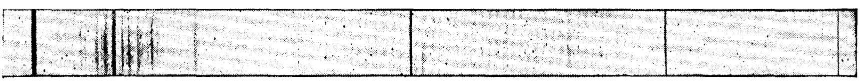


FIG. 34–15

4, 5, etc., the formula yields the visible hydrogen lines in order, beginning with the red and continuing out into the ultra-violet. This series is usually known as the Balmer [2] series. If higher values for n are chosen, several series of lines in the ultra-red are derived. The simplicity of this formula fits in perfectly with our picture of the hydrogen atom as the lightest possible one, consisting only of a nucleus and a single electron. It required, however, some very novel and radical assumptions, first made by Einstein [3] and Bohr [4] (p. 545) to furnish an "explanation" of even this simplest of all spectra.

[1] Named after J. R. Rydberg (1854–1919), professor of physics at the University of Lund, Sweden. He first found the approximate law (p. 543) which gives the distribution of the lines in spectrum series other than hydrogen.

[2] J. J. Balmer, a teacher in Basel, Switzerland, (1884), who was the first to solve this puzzle and find the law connecting the wave-lengths of this series.

[3] A. Einstein, of Berlin; perhaps the most imaginative of living mathematical physicists; originator of "Einstein theories" on a wide variety of subjects. The one which has made his name almost a household word is his great work on relativity, but his other contributions to modern physics are of equal interest.

[4] Niels Bohr, Professor in the University at Copenhagen.

Other spectra, such as those of helium, sodium, calcium, aluminum, carbon, etc., consist of many series, in some of which the lines are single, in some double (we have already noted the pairs of lines in the sodium spectrum), in some triple, and in many in the form of little groups of greater or less complexity called in general ***multiplets.*** The series in most of these cases follow approximately an empirical formula of the form

$$\frac{1}{\lambda} = L - \frac{R}{(m + a)^2}$$

known as Rydberg's formula, in which L is the limit of the series, expressed in reciprocal wave-lengths, or "wave-number" units,[1] R is the Rydberg constant, whose value is practically constant for all series in all elements, m is any integer, and a is a constant peculiar to the series in question. This formula is of less interest than might be, for the reason that it is not exactly followed by the lines of the series, especially for small values of m; but it is given here to indicate the manner in which the simplicity exhibited by hydrogen breaks down in passing to more complex elements. It is now believed that the spectra of ***all*** the elements can be analyzed into groups of series, with no lines left out of the scheme. Indeed, the formidable task of disentangling these complexities has already been accomplished for a large number of spectra.

Figure 34–16*a* shows a couple of series of triplets in the ultra-violet part of the spectrum of magnesium, running to a common limit. Figure 34–16*b* shows two series of close pairs in the visible part of the sodium spectrum. Figure 34–16*c* shows part of the helium spectrum in the ultra-violet in which members of six series appear, marked by distinguishing characters. In each case a scale of wave-lengths is drawn above.

Bohr's stationary states. Bohr in 1913 had the courage to propose a bold theory of atomic structure which has since led to great advances in our understanding of atoms, and in particular has helped to solve the puzzle of the nature of line spectra. He imagined the electron in the hydrogen atom as occupying any one of a series of approximately circular orbits, in which it revolved about the nucleus, each orbit being of a different size and involving a different amount of energy. He endowed the electron with the ability to continue revolving in one of these orbits for an appreci-

[1] The number of waves per centimeter length *in vacuo* is usually chosen.

able time without loss of energy, contrary to the ordinary laws of electrodynamics, according to which it would be expected to radiate its energy away in the form of electric waves and fall in on a spiral path toward the nucleus. Thus, while the electron is in one such orbit, the atom as a whole is supposed to have a certain constant

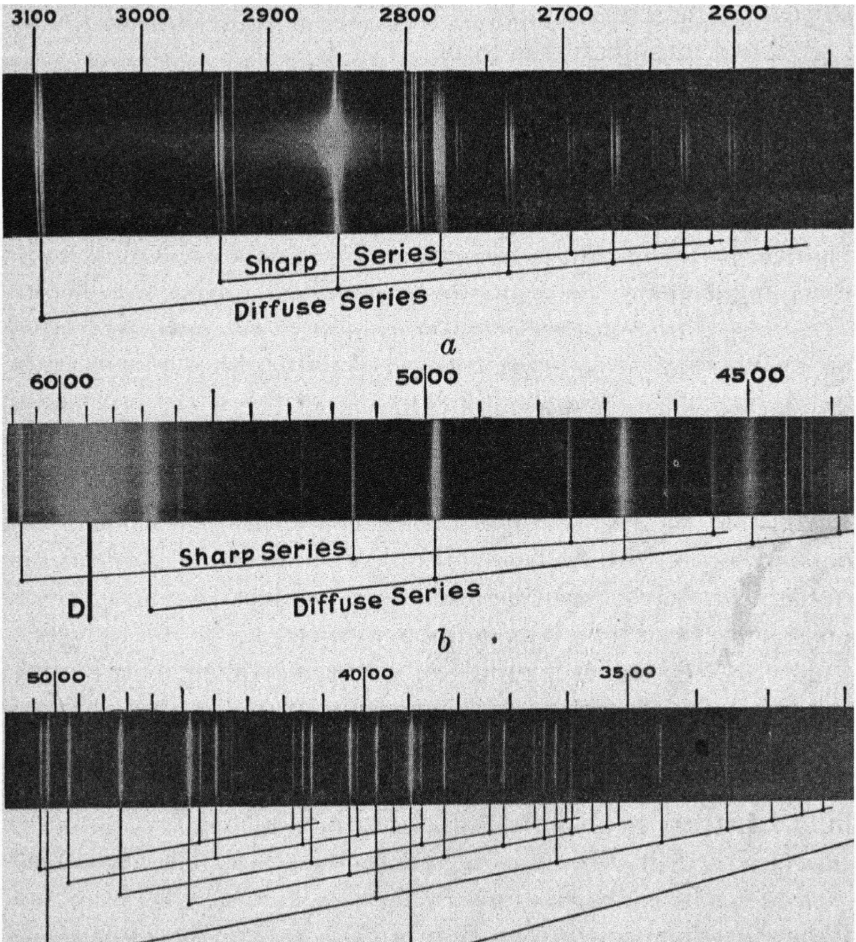


FIG. 34-16

Emission spectra showing series

amount of energy, and is said to be in a ***stationary state.*** When, in Bohr's picture, the electron changes suddenly from one orbit to another, the atom as a whole changes its energy and passes from one stationary state to another, radiating out the difference of energy in the form of light of one wave-length only. Later investigations have shown that we may perhaps have to abandon Bohr's

picture of orbits and revolving electrons, but we may well retain the idea of stationary states of the atom as a whole, without attempting to form a more definite picture of their formation than to say that they are due to alterations in the position or state of the electrons.

The emission of light in quanta. When an atom loses energy by passing from one stationary state to another, we cannot suppose this energy to be destroyed, that is, so long as we adhere to the principle of the conservation of energy. It is either passed to another atom through a collision or radiated out in the form of heat or light. We have no picture as to how this can happen; in fact, in this theory we must learn to get along as best we can without pictures or mechanical models. Especially is this true in regard to the emission process. Bohr adopted the idea, originated by Planck [1] in a theory explaining the form of the curves of energy emitted by black bodies, Fig. 34–12 (p. 535), that ***energy can be emitted or absorbed only in small but definite amounts, called quanta or photons,*** and that the change of energy from an amount E_1 to amount E_2 is given by the equation

$$E_1 - E_2 = hf,$$

where f is the frequency [2] of the wave sent out and h is a universal constant now named after Planck. Such an equation was first devised by Einstein in a different connection (p. 598). These quanta need not all be of the same size, as atoms are supposed to be; but each size of quantum is emitted by means of waves of a different frequency, a small quantum giving rise to a low frequency, and a long wave-length. To show how very unmechanical this idea is, one might imagine a ball rolling down a flight of unequal steps and emitting the energy lost at each step by sending out a series of sound vibrations. The frequency of these vibrations would have to be proportional to the height of the step through which the ball has fallen. Of course, there is no way of persuading a ball to do this. If it does send out energy in the form of sound waves, it sends out the same sort of waves for all heights of step,

[1] Max Planck, professor in the University of Berlin.

[2] Uniformity (in this book) suggests the use of the letter f, though the Greek letter ν ("nu") is more often written for the frequency in this equation. This letter, however, besides being confusing, is also commonly used for the number of waves per centimeter ($1/\lambda$).

but louder ones for greater heights. The atom, on the contrary, sends out higher frequencies for higher energy-steps.

Bohr's picture of the cause of the hydrogen spectrum. If the energies of the hydrogen atom in different stationary states are represented as steps in an unequal ladder, down which the single active electron may fall, we can then account simply and quantitatively for the main features of the hydrogen spectrum. Thus, if a normal hydrogen atom experiences some sort of accident whereby it loses its electron and later recaptures it, or another one like it, the home-coming electron may be thought of as dropping in from outside down this ladder, jumping from one level to another, not necessarily touching every one. The jump from the second (counting from the bottom) down to the first or normal-atom level (Fig. 34–17) causes the emission of the first line of Lyman's series, whose wave-length is 0.00001216 centimeter, or 1216 Ångström units [1] ("A.U."). The jump from the third level to the second involves less energy and therefore gives the red line (α) of the Balmer series at 6563 A.U. The jump from the third to the first yields the second line in the Lyman series; from the fourth to the second, the second line (β) of the Balmer series; from the fourth to the third, the first line of an ultra-red series, and so on. Once the fundamental

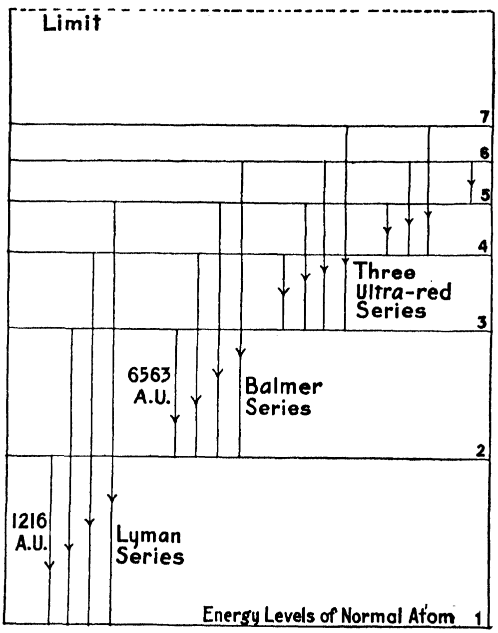


Fig. 34–17
Energy levels of the stationary states of the hydrogen atom

[1] One A.U. = 10^{-8} cm. Wave-lengths are usually measured in terms of these units, named in honor of A. J. Ångström (1814–1874), professor of physics at Upsala University, whose work furnished a foundation for spectrum analysis, and who discovered the presence of many familiar substances in the sun.

hypotheses of this theory are admitted (one might almost say ***swallowed***), the rest is easy, and so much that is true follows from these hypotheses that they cannot be ignored.

The derivation from Bohr's theory of the hydrogen series formula with the ***correct numerical value*** for the Rydberg constant is one of the triumphs of the theory. It is not difficult, but is omitted here. The result is that

$$R = \frac{2\pi^2 me^4}{ch^3};$$

it is thus seen to depend on the charge of the electron, on its mass m, the velocity of light c, and Planck's constant h.

Extension of Bohr's theory to other atoms. When an atom contains many electrons, the inner ones are believed to have very little to do with the emission of its spectrum or with its chemical properties, both of which are ascribed to its outer electrons. These by their mutual forces affect the energy values of the atoms, and the number of possible stationary states is greatly increased (compared with hydrogen), so that we must now represent them by many ladders of energy levels. A recaptured electron may then return to its normal position by very many different routes; this process is accompanied by the emission of a large number of different spectrum lines, one at a time. The entire spectrum is emitted at every instant because there are always millions of atoms active and going through these changes. By no means all possible ladder-jumps seem to be allowable. Curiously enough, jumps down one ladder are rare; instead they are made from one level to another on an ***adjacent*** ladder. So many energy changes are available that this theory accounts without difficulty for the large numbers of spectrum lines actually observed.

In the case of many elements each energy level is multiple. In iron, for instance, threefold, fivefold, and sevenfold levels occur. Energy changes from one of these multiple levels to another give rise to multiplets, perhaps with fifteen lines or more in each instead of the single line hitherto supposed to arise from such a change. Thus the spectrum becomes a very rich one, the scheme of whose structure is much more difficult to discover.

Excited and ionized atoms. When an atom is in its normal state and one of its electrons is displaced by a collision or otherwise, so that the atom reaches one of its other stationary states, it

is then called an "excited" atom. In this condition it may emit light when it returns to the normal state. In general it will do so unless it collides in a particular manner (a "collision of the second kind") so as to hand its excess of energy over to some other atom without radiation. The emission of light is thought of as due to the return of excited atoms to their normal states. Atoms may become excited by being struck by other atoms which are thermally agitated, or by being struck by electrons which are moving fast enough to contribute a quantum of energy of the right size, or by absorbing a quantum of light emitted by some other atom. If they become excited to relatively low levels only, the light they subsequently emit is only a part of their ordinary spectrum, that part due to energy changes among the low-lying levels, and composed usually of the most important lines of the spectrum. This is what happens to the sodium atom in a Bunsen flame.

If an atom loses an electron altogether, it is then ***ionized*** and if it subsequently recaptures an electron, it may emit ***any*** line in the spectrum, and a mass of such atoms gives out the complete spectrum. The ***ionization potential*** is the potential difference through which an electron must fall in order to acquire speed enough to ionize the atom. Thus, an electron moving in a vacuum from one wire gauze "plate" to another between which is maintained a difference of potential of 5.12 volts may acquire speed enough to knock the outermost electron out of a sodium atom; therefore the ionization potential of sodium is said to be 5.12 volts. Such an electron has an actual speed of 1.3×10^6 meters per second, as may be determined by equating the electrical energy eE, spent in sending the electron through the difference of potential E, to the kinetic energy $\frac{1}{2}mv^2$ acquired by the electron on this account.

The ionization potentials of the elements measure, in a sense, the solidity or firmness of their atomic structures, just as one might indicate the strength of a building by specifying how fast a cannon ball would have to move in order to knock it to pieces.

Ionized atoms and their spectra. If an atom loses an electron which it does not immediately recapture, it becomes temporarily a quite different sort of atom. Of course, its positive and negative charges no longer balance, so that it is really a positive ion. Its nucleus is unchanged, and this means that it weighs as much as usual, the lost electron being insignificant in weight compared with the rest of the atom. But the chemical properties of this atom, and

the spectra which it emits if one of its remaining electrons is disturbed, (so that the atom becomes "excited,") are those of the element which normally has one less electron always. For example, aluminum has thirteen electrons. Two of these are close to the nucleus, eight more are arranged in the form of a symmetrical "shell" around the inner group, and three are left over as external electrons. If one of these is removed, the arrangement of the electrons that remain becomes very like that of magnesium, which normally has twelve electrons. If two of the three external electrons of aluminum are removed, the remainder take an arrangement like that of sodium, which possesses a normal quota of eleven in all. This sodium-like atom is called *twice-ionized aluminum* and denoted by the symbol Al^{++}.

It must not be concluded that the spectrum of Al^{+} is *identical* with that of magnesium, or that of Al^{++} with sodium. The spectra are built on the same plan; all the lines are, for instance, doubled in sodium and in Al^{++}, or composed of two groups of singlets and triplets in magnesium and in Al^{+}. But the fact that there is an unbalanced charge on the nucleus in such ions makes an important difference; the forces of attraction holding the electrons are greater than usual; their energies in the different stationary states are greater, and so also are the differences of energy, which are proportional to the frequencies of the emitted lines. Hence the spectrum of Al^{+} as a whole is shifted toward the higher frequencies or shorter wave-lengths (i.e., into the ultra-violet), as compared with that of magnesium, and that of Al^{++} is farther yet.

If the chemist could only gather a considerable quantity of Al^{++}, he would have a new element with the properties of sodium but the weight of aluminum. Of course it would be so highly charged that it would pick up electrons easily and revert to ordinary aluminum. Thus, though these ions can be regarded as chemically new substances, we can never hope to perform chemical experiments with them. Their spectra, however, show us how they are built, and we can observe their brief careers in sources of light. As many as six different spectra have been observed to come from the oxygen atom when disturbed with greater and greater violence, the last of them being due to the removal of five of the six external electrons usual with that element, turning the oxygen temporarily into a lithium-like substance. If most of the ninety known elements can be taken apart in this manner and each made to produce several new, though temporary, substances, it is evident that research in this field will not soon come to an end.

This disruption of the atoms occurs to a certain extent in the electric arc but best in violent sparks, preferably *in vacuo*, such as are obtained by the sudden discharge of several large Leyden jars.

Books recommended:

A. Haas "Atomic theory," 1927 (Van Nostrand).

E. N. da C. Andrade "The Structure of the Atom," 3rd ed. 1926 (Harcourt, Brace & Co.).

H. A. Kramers and H. Holst, "The Atom and the Bohr Theory of its Structure," 1923 (A. A. Knopf).

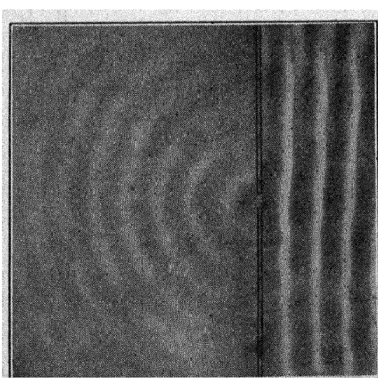
J. K. Robertson, "Introduction to Physical Optics," 1929 (Van Nostrand).

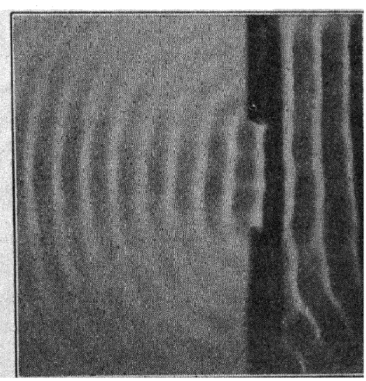
CHAPTER 35

DIFFRACTION, COLOR, AND INTERFERENCE

DIFFRACTION

The spreading of waves. We have already noted (p. 475) certain curious patterns in the shadows of common objects which are produced by the spreading of light waves around the edges of





the objects or apertures. This spreading is a phenomenon common to all waves, and is known as ***diffraction.*** It is more marked the longer the wave is, or the smaller the obstacle or opening. In Fig. 35–1, the spreading of short water waves (leftwards) into the quiet space beyond an opening in an obstacle (a sheet of metal) is shown when the opening is (*a*) narrow compared with the wave-length,

and (*b*) wide. The narrow opening allows a single Huygens wavelet (p. 481) to go through, with marked spreading on the far side, as required by Huygens' principle. The larger opening allows a section of the waves to pass through unchanged, spreading a little at the edges only. This phenomenon has many interesting optical consequences.[1]

Resolving power of optical instruments. The ability of any optical instrument to produce images which are accurate reproductions of the object in all its fine details is said to depend upon its ***resolving power.*** For instance, a lens of high resolving power is capable of producing separate images of two close objects, such as the two dots of a colon (:) as seen from three feet or more away. Omitting the mathematical definition of this quantity, we may still consider its physical significance. Any light wave passing through the limiting aperture of an instrument (e.g., the clear opening of the objective lens of a telescope) will be refracted by the lenses and made to form the image, excepting that near the edge of the opening there will be some straying of the light through diffraction, as in Fig. 35–1*b*. The proportion of this wayward light to that which goes where it is expected to varies with the size of the opening. The result is that one cannot see anything clearly through a ***very*** small hole (as one may verify by looking through a chink between two of his fingers), and even with a large opening, there are defects in the image, evident with high magnifying powers, which go with that size of opening and cannot be removed by any refinements in the optical system whatever. The result is that the image of a distant point source is rendered by a telescope as a small disc, fading away at its edge and surrounded by faint rings, seen greatly magnified in Fig. 35–2. A large opening makes the disc and rings small. In order to be able to see fine details in an image, these discs must be small compared with the details desired.

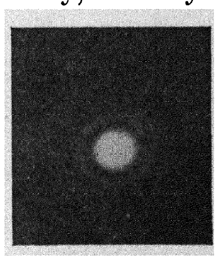


FIG. 35–2
The image of an artificial star taken with a small lens

The diffraction grating. A grating is composed of a series of

[1] These waves were generated in a small dish by means of vibrations created by the 60-cycle lighting current, and were projected greatly magnified upon a screen by intermittent light, also of 60 flashes per second, which then made the waves appear to be stationary.

bars with openings between. Such a thing may be made on a small scale, as Fraunhofer did, by stretching extremely thin wires over the threads of two parallel screws, or by ruling clear lines, regularly spaced, in a silver film deposited on glass. Still finer gratings can be made by photographing coarse gratings on a reduced scale. The best gratings are made by ruling fine lines on a reflecting metal surface (p. 554). Such an instrument is much used in scientific work to disperse the colors in white light into a spectrum. The reason why this happens and the difference between a prism spectrum and a grating spectrum must now be considered.

In Fig. 35–3, suppose that plane waves W are advancing upon the surface of a grating, here very much magnified. These waves will spread out beyond the openings at $A, B, C,$ etc., in ***all*** directions, if the openings are small compared with the wave-length, as in Fig. 35–1*a*. To show all directions at once in the figure would make it too complicated, but one direction is drawn, making an angle θ (the "diffraction angle") with the direction of motion of the incident waves.

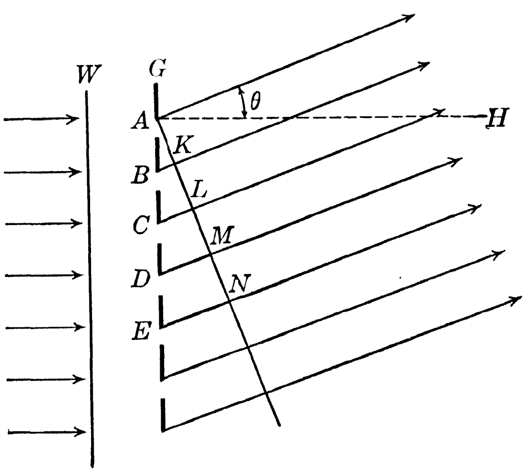


FIG. 35–3
The action of a grating

In this direction (as in every other also) there will be Huygens wavelets traveling from each opening. On the whole, the wavelet from B will be behind that from A by a distance BK, if AK is drawn perpendicularly to the diffracted rays. Likewise, the wavelet from C will be behind that from A by a distance CL, which is twice BK, if the spacings of the grating are regular. The wavelet from D will be behind that from A by three times BK, and so forth. If BK should happen to be just one wave-length (assuming the incident light to consist of one wave-length only) or should be equal to a whole number of wave-lengths, then all the wavelets would be in phase with one another along such a line as AM, and this would then become a new wave-front.

If a lens were to gather up these parallel rays (no lens is shown, as on this scale it would be huge) and bring them together at its focus, there would be at this spot a concentration of light, but at adjacent spots there would be no light, since for other directions the wavelets could not be all in the same phase and would therefore destroy one another by interference. To express these facts mathematically, let us call AB ($= BC = CD$, etc.) by the letter d, the "grating space." We may note that the angle BAK is equal to θ (since either angle when added to the angle KAH makes a right angle). The sine of the angle BAK is BK/AB. The condition giving the values of the angle θ at which the wavelets are in phase, and at which the concentrations of light of wave-length λ are to be found becomes

$$BK = n\lambda, \quad \text{or} \quad d \sin \theta = n\lambda.$$

This is the simplest possible form of ***the grating equation,*** applicable only when the light falls perpendicularly upon the grating.

Use of the grating equation. This equation informs us at what angles the wave-length λ will be found. The number n is an integer. If $n = 0$, θ must be zero; the equation is satisfied for all wave-lengths alike. That is, a white image is formed in a direction straight through the grating if white light falls upon it. If $n = 1$, $\sin \theta = \lambda/d$; that is, a given wave-length is found concentrated at this particular angle, on either side of the straight direction and at a distance which will be considerable if d is small, but which will become impossible if d is less than λ, since the sine of an angle cannot be greater than unity. The sine of the angle being proportional to λ with a particular grating (d constant), the wave-length can be measured easily and accurately by the simple operation of measuring θ and d. Indeed, d is a value often furnished with the grating by its maker, so that only one quantity need be measured. This is the commonest method of ***measuring the wave-length of light*** and furnishes results accurate enough for most scientific work, such as that on the nature of spectra which was discussed in the preceding chapter. Some of the values obtained will be found in Table XXX (p. 555).

The spectra formed by a grating. Figure 35–4 illustrates how spectra are formed by a grating. In the direction straight through there is an image for $n = 0$, the same for all colors. If a source of light is being used which emits a violet line near the edge of the

visible spectrum, it will fall at V_1 for $n = 1$, at V_2 for $n = 2$, and so on, approximately equally spaced unless the angles involved are very large. Similarly, a deep red line occurs at R_1, R_2, R_3, etc. White light forms a continuous spectrum between V_1 and R_1, another, called the ***spectrum of the second order,*** between V_2 and R_2, a third order between V_3 and R_3, and so on. It is evident that the spectra soon begin to overlap; thus between V_3 and R_2, there may be a mixture of red and violet lines in the visible spectrum, to say nothing of ultra-violet lines of the fourth and fifth orders, and ultra-red lines of the first order, all of which fall in the same region.

The whole set of spectra is repeated on the opposite side of the zero order image. The second order spectrum is twice as long as the first and usually much fainter; the third order is three times as long as the first and fainter yet. The higher orders are commonly too faint to be of use.

Comparison of prism and grating spectra. Either a prism or a grating may be used to form spectra. The prism economizes the light more than the grating, as it forms one spectrum only, but the grating arranges its spectra in proportion to their wave-lengths, and for measurements of wave-length this is often a great advantage. In a prism spectrum the red end is unduly compressed as compared with a grating spectrum; at the violet end, on the contrary, the dispersion produced by a prism is often very great.

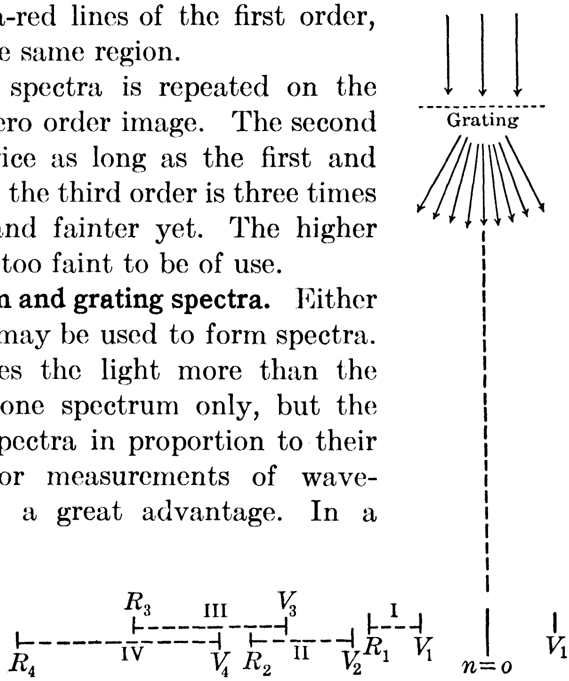


FIG. 35–4

Reflecting gratings and replicas. The best gratings for optical purposes are made by ruling fine equidistant grooves on a polished reflecting surface. Certain alloys make good reflectors, notably ***speculum metal,*** composed chiefly of copper and tin. A polished concave surface of this material will ***reflect and focus rays at the same time,*** and if it is ruled with 15,000 to 20,000 fine grooves to the inch, it makes a good ***concave grating,*** which needs no lens with it, but does its own focusing, as well as dispersing the colors. ***Plane gratings*** (of metal) are also much used, with lenses. A

plane grating may be covered with a thin film of transparent celluloid which dries and this coating can then be peeled off and mounted on a flat piece of glass. This makes a very convenient and common form of grating called a ***replica,*** which has ridges instead of grooves, and a surface which is all transparent. The replica operates, however, like the simple grating of bars and openings, because the thicker parts of the material delay the light waves and introduce differences of phase among the wavelets, upon which the action of the grating really depends. The transparency of the whole surface is useful because it increases the brightness of the spectra obtained.

Color and wave-length in the spectrum. With the diffraction grating accurate values of the wave-lengths of spectrum lines are obtained, and the ranges of wave-length included within each spectrum color are established. These are indicated in Table XXX. This table is extended to include other wave-lengths in comparison.

TABLE XXX

Wave-lengths of Light and other Electromagnetic Waves

Color	Approximate Range of Wave-length	Solar Fraunhofer Lines	Wave-length
		A	7661 A. U.
Red	7800–6300 Ångström Units	B	6867
		C	6563
Orange	6300–6000 " "	α	6276
Yellow	6000–5600 " "	D	5896 5890
		E	5270
Green	5600–4900 " "	b	5183 5173 5167
Blue	4900–4400 " "	F	4861
		G	4308
Violet	4400–3800 " "	g	4227
		H	3968
		K	3934

Radio waves	100,000 meters or more, to 0.3 mm.
Ultra-red	0.3 mm. to 0.00078 mm.
Visible	0.00078 to 0.00038 mm. or 7800 to 3800 A. U.
Ultra-violet	3800 A. U. to 100 A. U.
Overlapping region	100 A. U. to 10 A. U.
X-rays	10 A. U. to 0.1 A. U.
Gamma rays	1 A. U. to 0.01 A. U.
Cosmic rays	to 0.001 A. U. or further

The extreme ranges of the colors given in Table XXX are dependent on the brightness of the source and the condition of the observer's eye. Ordinarily the visible spectrum stops about 4000 A. U., but with the most intense sources it is possible to see below 3600. At the other end the limit ranges from 7000 to 8000. Similarly, the boundaries of the individual spectrum colors are somewhat variable.

COLOR

Colors produced by mixing lights. If two or more spectrum colors are thrown together on the same screen, new colors are obtained which are not seen in the spectrum itself. For instance, red and blue give purple, which is not a spectrum color. An interesting way of showing these mixed colors is to set up a simple one-lens form of spectroscope, as in Fig. 35–5. The light from a

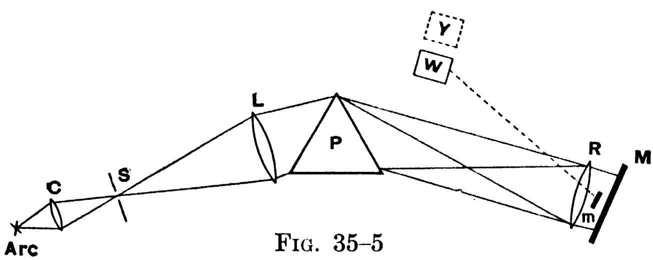


FIG. 35–5

carbon arc is concentrated by a condensing lens C on the slit S, and then falls on the lens L, which is convergent enough to produce an image of S some two or three feet beyond. A prism P is then interposed and a spectrum is formed at R. A very large lens can be placed at R so as to receive ***the whole spectrum on its surface*** and gather it all up into a white image somewhere beyond. This will be an image of the face of the prism. It is better, however, to place a mirror M just beyond R to return the whole spectrum through the lens R and form the image at W, somewhere near P, where it may be received on a screen.

The whiteness of the image W shows that ***the spectrum colors can be recombined into white light,*** which is just as it should be, since they were all derived from it. A more interesting result is obtained when a small strip of mirror (m in the figure) is held in front of M and in one of the spectrum colors. By tilting m slightly to one side, the light of the color in which it is placed can be made to form a colored image of the prism face at one side of W, say at Y. The light forming the image W will now no longer be white but will be a composite of all the colors which are left after m has

taken one away. The colors of the two images W and Y are ***complementary,*** that is, when they are added together, they form white, which can, of course, be done either by making m parallel to M, or by removing it altogether. As m may easily be shifted through the whole gamut of colors, there may be seen on the screen a succession of pairs of complementary colors. These may be varied still more by choosing a wider mirror for m so as to admit two adjacent colors at a time; or by having two mirrors like m, placed in different parts of the spectrum, so that three images are formed in all near W. These may be widely varied in color, and yet they always satisfy the condition that when superimposed they produce white light. The image at W is then always composed of the colors in the spectrum that have not been removed by the two mirrors m.

White light. What is meant by white light is somewhat difficult to define. Certainly straight sunlight is somewhat golden, and a north light on a clear day is bluish. The average light of a cloudy sky is presumably a mixture of sunlight and blue-sky color, which we may safely call the nearest approach to white light, and yet this will vary slightly in color with the height of the sun above the horizon.

The color disc. A disc may be arranged of three or more colored cards exposed in the form of sectors, as in Fig. 35–6, which may be spun by a motor at a high speed. The phenomenon of the persistence of vision in the eye enables one to "see" each strip of color for some time after it has gone by; so that at a high rate of spinning the colors become thoroughly mixed in the eye. If pure colors are available (they usually are not), a good grayish-white may be produced from red, green, and blue. These colors have been called the three ***primary colors*** for this reason and also because a fairly good match may be made for any other color by means of adjusted amounts of these three. Much depends, however, on having the right sorts of green and blue.

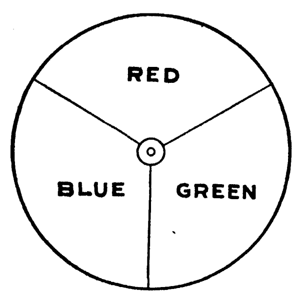


FIG. 35–6

When matching very mixed colors in this way, white and black cards may have to be added to the other three. If good permanent colors of definite chemical constitution are used, an exact specifi-

cation may then be made for any tint in commercial use, in terms of the angle of the sector in the disc which is occupied by each color. After these are known, the color can be correctly reproduced at any other place or time.

The colors of objects. A white object is one that sends back to the eye all colors equally well. It looks white in white light, red in red light and so on. A black object is one that reflects no light at all of any color. A red object is one which reflects red light, and looks red when either white or red light strikes it. If a red object is placed in pure blue light, however, it looks black. The colors of ordinary objects change with the kind of light falling on them. It makes a very striking experiment to darken the room thoroughly and then expose variously colored objects to the yellow light of a sodium flame. Since there is only one color present, all the objects exposed to this light have the choice of looking yellow or black, but there is no other alternative. A magical transformation occurs if white light is suddenly admitted, as by switching on an electric light.

Colors that are not identical may match well enough by lamplight, but not at all well by day. This is because daylight is richer in blue rays than the usual lamplight. In recent years "daylight" lamps of various sorts have been produced, many of which are tungsten incandescent lamps with a blue glass covering which cuts down the red and yellow light from the hot wire so that it bears the same proportion to the blue as occurs in daylight.

The colors of ordinary objects are due to absorption. The light going through a piece of red glass does so without material loss if it is red, but is absorbed if it is blue. The action of a piece of red blotting paper is similar. Since the "surface" is really an ill-defined and irregular mass of transparent red fibers, the light penetrates to a considerable depth in these and before it gets back to the eye of the observer, it has had to pass through some of the fibers. Thus, whether the paper is viewed by reflected or transmitted light, the color is the same, and always that produced by its absorption.

Colors produced by mixing pigments. If one glass transmits red light only, and another green only, the two together will transmit nothing, i.e., seem black. The absorptive powers of the two glasses are added in such a combination. Similarly, a blue pigment absorbs red and reflects green and blue; a yellow pig-

ment absorbs blue and reflects green and red; the two pigments mixed together have their absorptions added, i.e., they absorb red and blue, but they reflect green. This is, of course, the way an artist makes his green colors.

Colors in nature. In rare cases there is a selective action at the surface of a body which is well illustrated by a thin film of gold. Gold foil has a very definite surface (unlike the blotting paper already mentioned). Light which falls on it may do one of three things. It may be reflected, or it may enter and be absorbed, or it may be allowed to pass through. One can see through thin gold leaf, and the color is green; yellow and red are more or less "selectively reflected" by it; blue must be absorbed. This sort of "metallic" or "surface" color is seen in certain brilliantly-colored insects and in crystals of some organic compounds.

The color of the blue sky is explained elsewhere (p. 572). It is interesting to note that the blue colors which are found in birds' feathers (where not "metallic") are due to the same cause as the color of the sky; the bluebird's back, for instance, has no blue dyestuff in it, but there is in the translucent framework of the feathers a finely divided mass of particles which scatter blue light best. On the other hand feathers colored red or yellow contain pigments, while green is produced by a combination of scattering and pigment.

Interference effects such as occur with thin films (p. 562) are also found in the colors of insects and in feathers. These are usually produced by layers of a transparent substance regularly laid down over one another, though there are many variations.

Color photography. Many processes for producing photographs in their natural colors are in commercial use. The "Agfa" and "Lumiere" processes are very successful and are somewhat alike. In these a glass plate is dusted with a thin layer composed of a mixture of transparent red, green, and blue particles, evenly distributed side by side, so as to produce a uniform gray color (a combination of "lights" as in the color disc). This layer is then coated over with a photographic emulsion sensitive to all the visible colors (panchromatic). The plate thus made is then placed in the camera with its glass side toward the lens. A red portion of the image falling on the plate sends light through the red particles, but it can go through no others. Behind each red particle there will be a black silver deposit after the plate is developed. The plate is then chemically "reversed" so that the places formerly black are now clear, and *vice versa.* Hence, when the plate is finished, white light falling on the place where there should be a red image finds a clear path through the red particles, but nowhere else; hence that part of the image is seen as red. Mixed colors

are reproduced correctly, since their light is passed partly by each of the three primary particles in proper proportion.

Colored motion pictures. Many processes for producing motion pictures in color are in use. In the simplest of these only two colors are used, a light red and a bluish green; though no blue is present it is surprising how little it is missed. In an early method pictures were taken in succession through alternating red and green screens, and projected similarly. Such films are without color themselves. The speed of projection is such that the successive colors are mixed in the eye through the persistence of vision.

In a better two-color method (the Technicolor process) the light is split into two beams and *simultaneous* pictures are made through red and green filters. These lie alternately on the film, and are later printed off on separate films of an odd sort, yielding not an ordinary black and white image but a

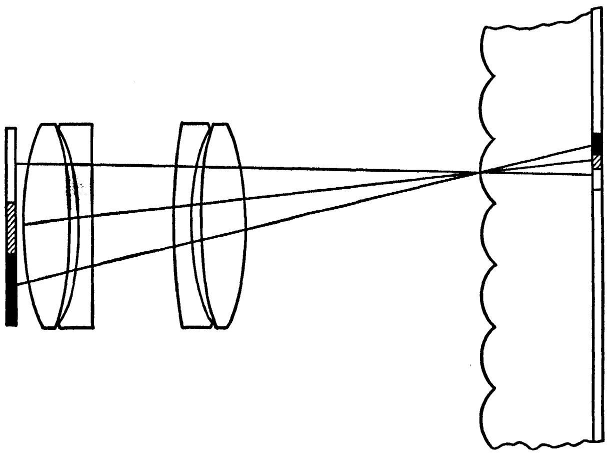


FIG. 35–7

Illustrating the process of colored moving pictures. (Courtesy of Eastman Kodak Company)

gelatin relief picture, the gelatin being thick where the image should be dark, and thin where it is bright. These are then dyed, and pressed in succession against a third film which absorbs some of their color. Thus the finished picture is formed like ordinary colored prints on paper, by the superposition and mixing of colors; in all such cases (p. 558) it is the absorptions of the two dyes which are added.

In the recent Eastman process for producing amateur motion pictures in color, three horizontal strips of red, green, and blue glass cover the camera lens. The rays of all colors are brought by this lens to the same focus and an ordinary black and white picture would result if an ordinary film were used. But this curious film is made of celluloid with a form, shown in Fig. 35–7, consisting of a series of cylindrical ridges with the sensitive photographic film at the back of the celluloid. These tiny cylindrical lenses are so strongly convergent that they form an image of the lens on the sensitive surface, so that every point in the picture is changed to three short parallel lines, minute

images of the three colored strips over the lens. If the object whose image comes at some one point is red, its light goes through the red screen but not the other two, and one tiny line, instead of three, is formed on the film. After development, and a special photographic reversal, white light can be sent through this part of the film in the opposite direction in the process of projection; this will pass through the clear spot where red light originally fell, and be refracted by the cylindrical lens so as to reach only that part of the projection lens where the red filter is placed, thus tracing out its former path backwards. Hence it comes out colored as it should be, and makes a red image on the screen. The film itself shows no trace of color, and the fine lines on it are so small that the images seem almost as sharp as usual.

Theories of color vision. Several theories to explain how colors are perceived by the eye have been proposed in the last two centuries. One of the most famous of these is based on the supposition that the eye contains three sets of apparatus, one for the perception of each of the primary colors, red, green, and blue. This theory assumes that the yellow sensation, for instance, is caused by the simultaneous stimulation of the red and green apparatus. It explains some of the facts of color blindness, but is not entirely satisfactory. Some recent observations have established the fact that when the left eye sees one color and the right a different one, and the two are made to overlap in one's field of view, a mixture color is seen where they do so. This mixture cannot be formed in either eye. It would therefore seem to be necessary to assume that part at least of our apparatus for color perception is located in the brain. There are many other difficulties. Physicists gladly leave this complicated subject to the psychologists and the physiologists.

INTERFERENCE

Interference in thin films. The word "interference" is used in optics to describe the combination of two or more waves in groups, so that their effects become additive. In the diffraction grating the wavelets from the separate openings interfere in this sense, and destroy one another except at certain angles where they agree in phase and act together. Such beautiful examples of interference as the colors of soap films, or of oil drops spreading over a wet pavement, are familiar to us all. The gorgeous play of colors which they exhibit challenges the attention and calls for an explanation, which the theory of the interference of light waves immediately supplies.

In these cases a wide beam of light, which for convenience we shall first suppose to consist of a single wave-length only, is being reflected from all parts of two surfaces. If we are examining the effects visible in a thin horizontal soap film, the light is reflected equally from the top and bottom surfaces. Such an incident ray

as AB (Fig. 35–8) will be partially reflected from the upper surface as BC, but most of it will be refracted along BD to the lower surface, where again a little of it will be reflected along DE and will emerge along EF. About nine-tenths of the light will go on through the film. If rays parallel to AB are thought of as striking all over the upper surface, there will also be rays such as BC and EF coming off from all parts of the same surface, and under these circumstances interference can occur. The two sets of reflected beams are nearly equal in brightness but differ in phase because one has gone somewhat farther than the other. These differences in phase are not hard to calculate. Suppose, for example, that the incident ray AB (Fig. 35–9) sends off a reflected ray EF traveling

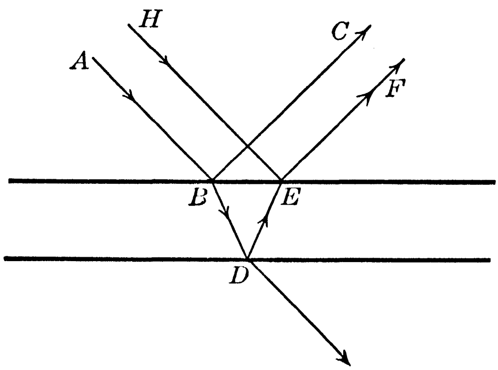


FIG. 35–8

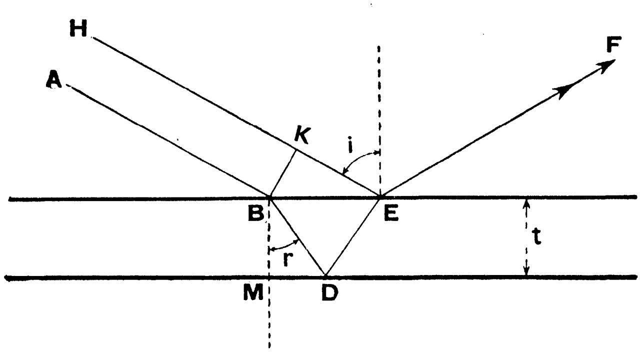


FIG. 35–9

via BD and DE, and that the parallel ray HE is reflected directly along EF. These two will then interfere along EF, and one will be behind the other by a certain amount. When this amount is calculated [1] the result shows that destructive interference occurs

[1] One ray travels through the film a distance from B of $BD + DE = 2 \times BD$; this is equivalent to n times as great a distance in air (n being the index of refraction), as far as delay is concerned, since the light travels n times faster in air than in the film; hence $2n \times BD$ is the equivalent air path. The

when a certain relation holds between the wave-length, thickness, index of refraction, and angle of incidence (or of refraction). With thin films when one wave-length is thus obliterated, others of nearly the same length are greatly weakened, so that when white light falls on the film a wide stretch of wave-lengths is practically cut out, others are partially reduced, and still others are reënforced and strongly reflected. The resulting color is brilliant and is made up of a mixture of those that are reflected. Pure spectrum colors never occur in such cases; the colors of thin films are always mixtures, "complementary" (p. 557) to those that are cut out. When the thickness is considerable, a large number of wave-lengths regularly spaced along the spectrum are destroyed at once, and yet so many different ones remain in between that the mixture makes white light, indistinguishable from the incident light, unless it is analyzed by means of a spectroscope. For intermediate thicknesses complex color mixtures are obtained in which pink and a light whitish-green are prominent.

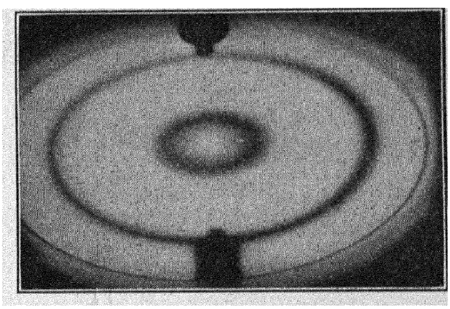


FIG. 35-10

A horizontal soap film in rotation showing two rings along which the reflected light is destroyed by interference.

All these interference effects should be viewed by reflected light. When one looks through a brilliantly-colored soap film, very little color is seen because interference then occurs between two beams of very unequal intensity. The "rainbow cup" offers one of the best means of studying such colors as are given by soap films. With it a hori-

other ray in the same time travels over a distance $KE = BE \times \sin KBE = 2MD \sin i = 2BD \sin r \sin i$ in air. Since $\sin i = n \sin r$, the difference in path by the two routes is $2nBD - 2nBD \sin^2 r = 2nt \cos r$, t being the thickness of the film, BM. There is a gain of half a wave-length at the reflection at B in air against the denser medium, in analogy to the reversal of phase in sound reflection at the fixed end of a rope (p. 254). If destructive interference is to occur, the two waves must differ by an odd number of half wave-lengths, or $(2N - 1)\frac{\lambda}{2}$ must be equal to $2nt \cos r - \frac{\lambda}{2}$, where N is an integer. Hence $N\lambda = nt \cos r = t\sqrt{n^2 - \sin^2 i}$, which gives the condition for no reflected light of wave-length λ. If the film is of air between two flat glass plates, the corresponding formula is $N\lambda = t \cos i$.

zontal circular film (Fig. 35–10) may be set into rotation about its center in its own plane, so that centrifugal force thins it out gradually, and the successive changes of color with increasing thinness can readily be projected on the screen, if a slowly convergent arc-light beam is reflected from the film to the projection lens.

A vertical soap film enclosed in a water-saturated space will gradually be thinned out at the top by gravity and exhibit the same colors. Figure 35–11 shows the aspect of such a film when viewed by light of one color, obtained most readily from a sodium flame (photographed with a yellow color screen to cut out the violet light from the flame itself).

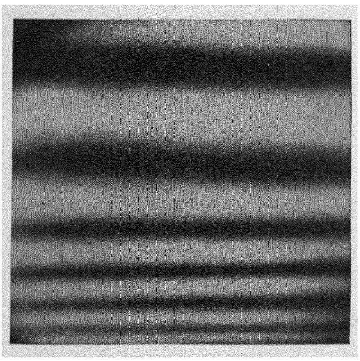


Fig. 35–11
Horizontal interference bands in a vertical soap film

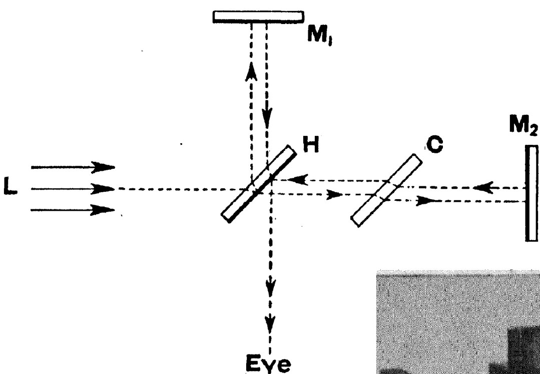
Practical uses of interference in thin films. In tempering a steel tool the mechanic notes the color which heat produces on the polished steel surface. A straw-yellow color signifies that a thin film of oxide has been formed, and that there is interference for the violet rays from the upper and lower surfaces of this film. The thickness of the film is indicated by the color, and in its turn this indicates the temperature to which the metal has been raised.

If an optician is making a flat glass surface, he may test the progress of his work by laying the new surface on a true flat one, and viewing the air film between them by reflected sodium light from a flame. If the surface is irregular, the variations in thickness will be disclosed by the forms taken by the dark interference bands seen in the air film. It is easy in this way to detect changes in thickness of a fraction (say one-tenth) of a wave-length of light (or 1/500,000 inch). Such high precision as this is needed in the most refined optical work, though ordinary mechanical jobs do not require it. Nevertheless, interference methods of this sort have sometimes been used to establish standards of length for mechanical processes (e.g., standardizing the diameter of shells for large guns, or the cylinders of automobiles).

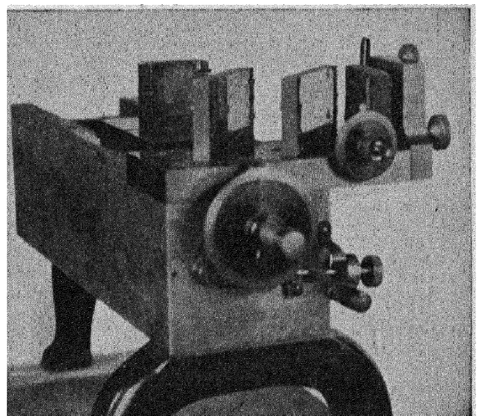
The interferometer. Such an uncanny delicacy of perception and measurement as has just been mentioned was brought within the reach of the laboratory physicist largely through the skill and ingenuity of Michelson, by his inven-

tion of the *interferometer*. This instrument enables one to measure ordinary lengths, not in terms of inches or centimeters, but in terms of wave-lengths of light as units. Its construction is simple, though the instrument itself must be very accurately finished.

In this instrument (Fig. 35–12*a*) a beam of nearly parallel light, L, is directed, perhaps from a sodium flame, through a lens toward a plane-parallel piece of glass, H, set at an angle of 45°. The side of H which is shown as a heavy line is "half-silvered," that is, coated with a film of silver (or nickel or platinum) so thin that about half the light falling on it goes through, the rest being reflected. The beam thus divides into two parts, one of which goes up to the completely silvered mirror M_1 and returns to H again, the second going to M_2 and back. The first has to travel through the plate H three times to reach the eye as indicated in the figure, and if we wish to make the two light paths identical, a second plane-parallel piece of glass, C, cut from the same piece as H, must be inserted on the way to M_2. One of the mirrors (M_1) is mounted so that it can be moved toward or away from H by turning a micrometer screw with a divided head which is seen facing the reader in Fig. 35–12*b*. This is very accurately made and can move M_1 by very small but still well-measured distances.



(*a*) The light-paths in a Michelson interferometer



(*b*) The interferometer itself

FIG. 35–12

If the incident beam is parallel and the two light paths of exactly the same length, the divided beams on their return to H will go off (half of each, at least) together in the direction of the eye, or observing telescope. The observer will then see a bright and uniformly illuminated field of view. If the movable mirror is then backed away through a distance of a quarter of a wave-length, the field of view will darken to a minimum of brightness which will be the more nearly black the more exactly *half-silvered* the plate H is. The light path on the side of M_1 has been increased by a half-wave in all (over and back) and thus the two beams are in opposite phase and interfere with each other. When M_1 has been moved a half wave-length, brightness is again restored.

It is almost impossible, and not very desirable, to secure such a condition of uniformity over the whole field of view. Usually one of the mirrors is tipped, in which case the field of view is crossed by *fringes* alternating bright and dark. Then, when M_1 moves, these fringes travel across the field of view. If the observing telescope has cross-hairs, the moving fringes can be counted. Thus if 100 fringes have gone by, the distance through which M_1 has moved must have been 50 wave-lengths. Distances that are measurable with ordinary precision by the micrometer screw which moves M_1 are also directly measurable in terms of light waves. A distance of about one-millionth of an inch, or one five-hundred-thousandth of a centimeter can thus be observed.

Uses of the interferometer. Michelson applied his interferometer to two fundamental measurements, or rather to one measurement with two fundamental aspects, the determination of the number of wave-lengths of light in the standard meter, or the accurate measurement of the wave-lengths of several spectrum lines. From the first point of view we now know that if the standard meter is destroyed, we can reproduce it by comparison with light waves to an accuracy of about 1 part in 5,000,000. From the second point of view, we are able to measure the wave-lengths of spectrum lines 100 or 1000 times more accurately than can be done with the diffraction grating. This may almost seem to be an unnecessary step, but the history of science shows that the more accurate our methods of measurement become, the more discoveries of importance we are likely to make.

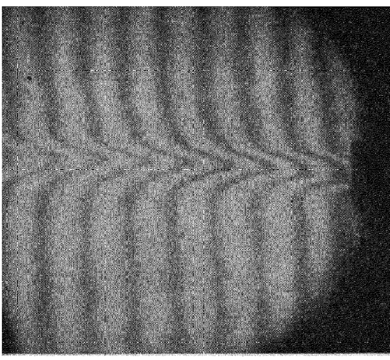


FIG. 35–13

The interferometer also makes it easy to measure very small motions, the expansion of a small crystal, for instance, when it is slightly warmed, or the heaving of the "solid" earth under the tidal actions of the moon and sun. It enables us to find the delay experienced by a beam of light in passing through a few inches of air as compared with an equal length of vacuum, and hence to obtain the index of refraction of air.[1] It is used in many sorts of physical research in which the utmost refinement of measurement is required. Figure 35–13 shows the fringes in the interferometer distorted by a small stream of carbon dioxide blowing leftwards out of the tip of a glass tube; from the maximum shift produced in the fringes, together with the thickness of the gas stream and the wave-length of the light, the index of refraction of the gas can be obtained.

[1] The index of refraction of any gas is defined as the ratio of the velocity of light *in vacuo* to the velocity in the gas. It is a number only very slightly above unity.

PROBLEMS

1. How is it that two colors that match in artificial light may prove to be different in daylight? What colors should prove to be the hardest to match correctly by artificial light?

2. If an artist had a pure yellow pigment (i.e., reflecting nothing but yellow) and a pure blue pigment, what would he get when he mixed them?

3. Why is a lake blue on a very clear day? Why, if it is rough, are the "white-caps" white?

4. Why does snow, which consists simply of small transparent ice crystals, look so unlike ice? Why is it not blue like water on a clear day?

5. A blue crystal, like copper sulphate, can be powdered, but is then white. Explain.

6. A grating could be ruled with 50,000 grooves per centimeter. What sort of spectra could be formed by it, and how would it differ in appearance to the eye from an ordinary grating, with perhaps 6000 grooves per centimeter?

7. A beam of parallel light of one wave-length only falls perpendicularly on a grating which has 7000 grooves per centimeter, and is diffracted out to a first-order image at an angle of 30°. What is the wave-length?

8. What color would be cut out of white light reflected perpendicularly from a soap film of thickness 0.00003 cm., and of index 1.34? What would be the color of the reflected light?

9. Figure 35–13 shows that a stream of carbon dioxide 2.1 mm. thick shifts the fringes of sodium light ($\lambda = 0.00006$ cm.) in an interferometer by 1.1 fringe. Remembering that the light passes twice through this stream of gas, how much path difference does this stream introduce, as compared with the same thickness of air? The difference in path varies with the difference in speed. Hence find the difference in index of refraction of the two gases, assuming that the index for a gas is the ratio of the speed of light in a vacuum to that in the gas.

10. What other wave-lengths occur at the same angle as 7000 A.U. in the third order of a grating spectrum? Which are visible, and what color are they? Why does the answer not depend on the width of the grating-space?

Books recommended:

R. A. Houstoun, "Light and Color," 1923. (Longmans, Green).
E. Edser, "Light for Students." (Macmillan).

CHAPTER 36

POLARIZED LIGHT

The nature of polarized light, 568; the production of polarized light by reflection, 569; Brewster's law, 570; color and polarization produced by scattering, 571; the blue sky, 572; polarization by double refraction, 572; the Nicol prism, 573; a polarization photometer, 574; mechanical explanation of double refraction, 575; uniaxial crystals and the wave surface, 576; the uses of the polariscope, 577; artificial double refraction, 578; rotation of the direction of vibration, 579; effects of magnetism on light, 580.

The nature of polarized light. It seems quite impossible for the ordinary human intellect to conceive of waves of light without imagining a medium in which these may occur (p. 178). Moreover we think of a wave motion as involving displacements of particles of the medium (in this case the "ether"), and these displacements must have directions also. In applying these notions to waves of light we may merely be exposing the weaknesses of our intellects, but at present we have no alternative.

Huygens thought of light waves as being due to longitudinal motions, just as sound waves are, but we have now to consider a series of experiments which make it seem quite clear that they must be transverse. We know that light waves, heat radiation, and electromagnetic waves can all be reflected, refracted, diffracted, dispersed, and, to anticipate a little, polarized. In electromagnetic waves (p. 434) the electric displacement is in one direction, the magnetic displacement at right angles, and both are in the plane of the wave-front, perpendicular (under all ordinary circumstances) to the direction of motion of the wave. Evidently, these arrangements hold unchanged for light waves. No complete history can be given here of the way in which the question of "the direction of the light vibrations" was settled, but some experiments connected with this question will be considered.

The term ***polarized light*** is used to describe that sort of light in which the vibrations in the wave-front are restricted to one particular form or direction (a "pole" is a direction). Thus, in ordinary polarized light (often called plane-polarized) the "particles" in

the wave-front are all doing the same thing at the same time, and this consists in vibrating back and forth through equal distances along parallel lines. The "direction of vibration" is any line parallel to these, and this is always taken to be the direction of the electric vibration in the wave-front, rather than the magnetic.[1] If each particle is describing an ellipse and all these ellipses are similar, we call this light ***elliptically polarized.*** If the particles are all describing circles, the light is said to be ***circularly polarized.***

The production of polarized light by reflection. There are three ways of producing polarized light. One is by reflection. When light falls upon a smooth glass surface at a moderate angle, four or five per cent of it may be reflected; the rest enters the glass and can be ignored for our present purposes. (Black glass is commonly used for this experiment, or ordinary glass painted over on the back with a coat of black "duco" varnish, so that there will be only one reflected ray.) If this reflected beam then falls upon another surface, the largest proportion of it is reflected there if the second reflection occurs in the same plane; that is, when the incident, the once-reflected, and the twice-reflected rays are all in the same plane (as in either of the arrangements shown in Fig. 36–1). If the second surface is then rotated through an angle of 90° about the once-reflected beam as an axis, and if the angles of incidence are varied, but kept always equal at the two surfaces, it is found that the second surface refuses to reflect the beam of light at one particular angle of incidence (57° for ordinary glass; more for substances of higher index of refraction), and then transmits it all, if the material is transparent. The angle of in-

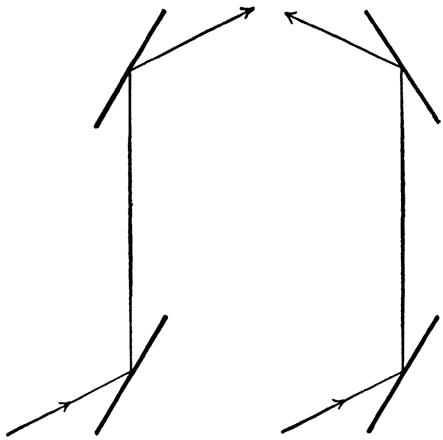


Fig. 36–1

[1] Electric waves produced by sparks between two conductors can be reflected by wire gratings only when the wires are parallel to the electric vibrations. Similar experiments, but on a much reduced scale, disclose the direction of the electric vibration in heat radiation and in light waves.

angles. Hence $i + r = 90°$, since ROP is itself a right angle; and $\sin r = \cos i$ (p. 620). Therefore the index of refraction n is now given by the equation

$$n = \frac{\sin i}{\sin r} = \frac{\sin i}{\cos i} = \tan i.$$

This is Brewster's law,[1] and it furnishes a useful way of finding indices of refraction of small specimens; or, conversely, of calculating polarizing angles, if the indices are known.

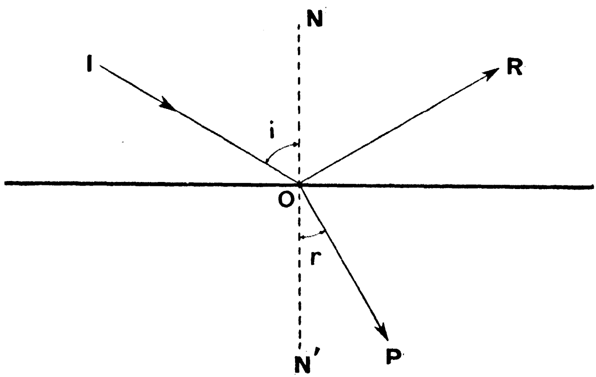


FIG. 36–2

Color and polarization produced by scattering. When light falls on particles as large as the water drops in clouds, it is reflected from each drop much as usual, and no peculiar effects are observed. When the particles are very much smaller, however, like those in the smoke arising from the burning end of a cigarette, the light is scattered rather than reflected regularly, and a distinct bluish tinge is to be seen in this scattered light. The particles are now probably somewhat smaller than the waves of light, and such experiments show that ***very small particles scatter short waves better than long ones.*** The same color can be seen if a strong beam of light is sent into a jar of water in which a very small quantity of soap has been dissolved.

Under the same conditions that give this bluish color, the light scattered out in any direction perpendicular to the incident beam is partially, or in some cases wholly polarized. An arc-light beam polarized by some suitable device (as by being reflected from a pile of plates or refracted through them) can be concentrated by means of a lens and sent horizontally through a dilute soap solution. If the light vibrations incident on the solution are horizontal, the light will be scattered vertically, but there will be almost none in a horizontal direction, except from specks of dirt, or bubbles, in the

[1] Sir David Brewster (1781–1868), Scotch experimental physicist, editor, and scientific writer; especially known for his work in light, including practical improvements in the optical systems of lighthouses.

water, which being large will always seem to glow brightly. If the direction of the incident vibration can be changed, which can be most easily done if a Nicol prism is used as a polarizing device (p. 573), it is interesting to see how the light which is scattered horizontally increases in amount as the vibrations become more nearly vertical. If one observer watches above the solution, while another is stationed to one side, both of them receiving rays scattered at right angles, one will declare the solution brightest when the other sees it dark.

There is a simple mechanical explanation of this effect. If the incident light is polarized, vibrating in a horizontal line, it produces horizontal motions in the particles, or perhaps in the electrons in these particles. These in turn send out waves of "scattered" light in all directions. At right angles to the incident light, where the observations are made, there can be no vibrations except in a vertical direction, because ***horizontal ones would be longitudinal*** and we know from a variety of considerations that longitudinal vibrations do not occur.

The blue sky. Particles of matter, however small they are, scatter light waves, though the amount of scattered light is very minute when the particles are molecular in size. If there are enough of them, however, the light scattered even by molecules should be perceptible, and it is scattering of this sort that sends us the light we receive from a clear sky. The clearer the air the bluer is the light of the sky, and this is reasonable, since particles as small as the air molecules ought to scatter short waves much better than long ones. Such light ought also to be polarized, and most conspicuously so in a direction perpendicular to the rays of the sun, according to the reasoning in the last paragraph. This proves to be the case. The polarization is almost complete in clear air; in dusty air the large particles which are present contribute some natural or unpolarized white light by regular reflection and the sky is not so blue. The miles of thickness of the air through which we look make the light which is scattered by the molecules visible even though the contribution of each one must be excessively small. If it were not for the atmosphere, the sky would be black, and the stars could be seen shining at any time in the day.

Polarization by double refraction. A curious effect is observed if a crystal of calcite (iceland spar or calcium carbonate) is laid

upon a printed page. Every letter is seen double (Fig. 36–3), and if the crystal is turned around on the paper, one image seems to remain at rest while the other, due to what is called the "extraordinary" ray, revolves around it. This is an example of ***double refraction.*** If two rays are obtained by sending a strong beam of light through such a crystal, forming on a screen two images of a small source, it is easy to test the two images for polarization, and it is then seen that each is polarized, and that the directions of vibration in the two cases are at right angles to each other. This is always the case in double refraction.

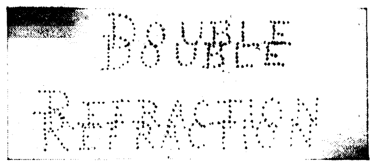


FIG. 36–3

Photograph (negative) taken through calcite. The letters were were made of pin holes in black paper.

Double refraction is observed in a great variety of crystals. The very simplest (cubic) types, such as rock salt, do not show it. Quartz and calcite are the most useful. Such materials have two ***indices of refraction,*** since the two rays have different angles of refraction and different velocities in the crystal.

The Nicol prism. Two pieces of calcite may be cut in a particular manner and put together as indicated in Fig. 36–4, cemented along the face CB with Canada balsam. If the crystals are properly cut, an incident ray I divides into two, O and E, on entering the crystal, which at first follow paths only slightly different. These then strike the layer of Canada balsam, and one (the ordinary ray) is totally reflected and lost, while the other goes on through the Nicol[1] "prism" and furnishes a beam of colorless, polarized light of nearly half the brightness of the incident one. The reason for this difference in behavior of the two beams at the surface CB depends on the indices of refraction. It happens that the index of the balsam (1.55) is intermediate between the indices of the calcite (1.66, 1.49). Thus the ordinary ray as it approaches the balsam is approaching a medium of lower index, and we have seen (p. 486) that

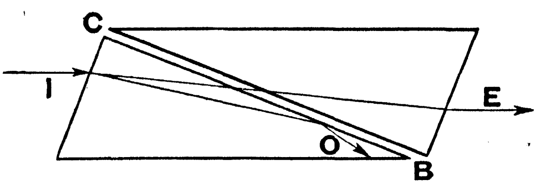


FIG. 36–4
A Nicol prism

[1] W. Nicol (1768–1851), Scotch physicist, whose name seems to be known chiefly by this invention (1828). He acquired great skill in working glass and crystal surfaces, and spent the latter part of his life as a recluse surrounded by his apparatus.

under these conditions, if the angle of incidence is large enough, total reflection will occur. A large angle of incidence is secured by the long slant surface between the two pieces of crystal. For the extraordinary ray, however, (index 1.49) the balsam is a medium of higher index, and the ray is bent into it, and so passes through.

The Nicol prism is the best polarizing device that requires no change of direction of the beam of light; it is colorless, and it can easily be obtained, at least in small sizes. If two are used in series, as in the *polariscope*, the first is called the *polarizer*, and the second the *analyzer*. While one can set up a cheap but efficient form of polariscope with a glass plate as polarizer and a small Nicol prism in the eyepiece, as in Fig. 36–5, a better form is made with two Nicol prisms. When one looks into a polariscope and turns one of the Nicol prisms, one sees variations in the brightness of the light getting through. When no light is able to pass through the combination, the prisms are said to be "crossed."

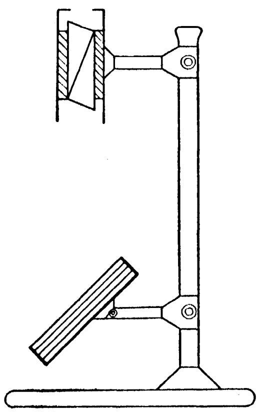


FIG. 36–5
A simple form of polariscope

These changes in brightness could not be explained if light were due to a longitudinal vibration. If in that case it could get through at all, it seems that it must do so equally well however either prism was turned. If, on the other hand, we suppose the light vibration to be transverse, it is easy to see that the structure of the crystal might affect the vibrations in a way which would alter as the crystal was turned. Hence from this sort of experiment as well as from the phenomena of scattered polarized light we conclude that *the vibrations of light are transverse.*

When one looks at the end of a Nicol prism, one sees a cross-section of the shape shown in Fig. 36–6. It is interesting to note that the "direction of vibration" (this expression always refers to the electric vibration) of the light is indicated by the arrows in the figure, i.e., along the narrow diagonal. This can easily be verified by sending a beam of light polarized by reflection into a Nicol prism and turning the latter so as to transmit the beam. The direction of vibration of a reflected ray is known (p. 570) and hence that of the transmitted beam is determined.

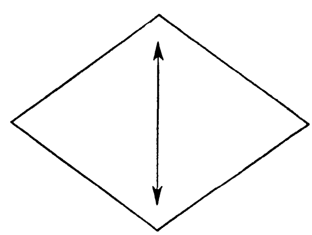


FIG. 36–6

A polarization photometer. There are many uses in scientific work for a device transmitting amounts of light which can be made to vary in a measurable manner. If light passes through two Nicol prisms in series, turned so as to transmit a maximum amount of light, the amount that gets through then may be taken as unity. If the analyzing prism is turned through an angle, an amplitude of vibration is transmitted which can be found, as in Fig. 36–7, by taking components. If AB represents the amplitude obtained when the

prisms are parallel (i.e., transmitting the maximum amount), CD represents the amplitude now obtained, which is $AB \times$ the cosine of the angle between. The energy of any vibration can be shown to be proportional to the square of its amplitude; $(AB)^2$ was taken as unity; hence at 45° one half the light gets through, at 60° one fourth. Thus a compact form of photometer can be made, the light through which can be varied in brightness by measured amounts, to match changes of some other beam of light. This devise is useful in measuring absorption, reflection from surfaces, etc.

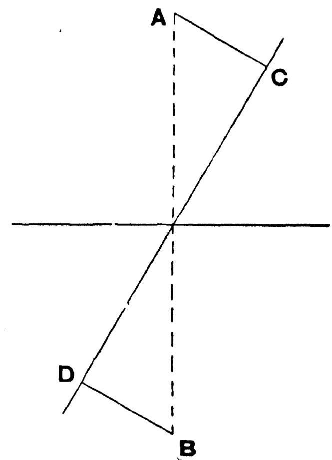


FIG. 36-7

Mechanical explanation of double refraction. A mechanical analogy may help to clear up some of the difficulties in this curious subject. Suppose that a rectangular bar of iron is obtained, two or three feet long, whose section and length are shown in Fig. 36-8. If this rests on two supports A and B, placed about a quarter of its length from its ends, it will vibrate persistently when struck with a rubber hammer in the middle (p. 237). These vibrations are due to the existence of elastic waves which travel rapidly from end to end of the bar and form standing waves in it. These waves travel because the bar is stiff and resists bending, and if it could be made stiffer, they would travel faster and the resulting vibrations would have a higher frequency. Thus the bar if laid down flat gives a certain low musical note, but when stood up on its narrow edge

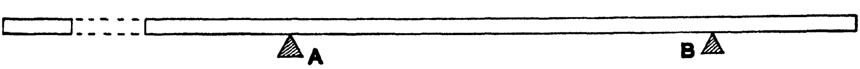


FIG. 36-8

gives a much higher note. We see that there are two velocities for elastic disturbances in this bar, which are quite different. A blow struck on the edge of the bar may make it emit two notes, but only two. This must be because ***two elastic waves are formed in the bar,*** with their two different velocities, but ***no intermediate velocities,*** or musical notes, ***are possible.*** A diagonally aimed disturbance automatically breaks up into two components, one of which travels with the slow speed and is due to displacements perpendicular to the wide face of the bar, while the other travels fast and is formed from displacements at right angles to the first.

If one had a bar of very great length, and struck its edge at one end, one of the two waves which would immediately start down the bar would gain on the other, and as the actual motion of any point on the bar must always be a combination of the two displacements, the actual form of the vibration of the particles would vary at different distances all along the bar. The corresponding optical idea is useful later (p. 578).

Thus it appears that one disturbance in such a bar breaks up into two which are "polarized," vibrating at right angles to each other, and traveling with different velocities. We could just as well have started with a square-sectioned bar of crystal (or even of wood) whose elastic properties are different in different directions, so that the variations in stiffness are due to these properties rather than to the shape of the cross-section. The result would have been the same. To return to a crystal through which light is traveling, we need only postulate such variations in elastic properties in different directions as would create differences in the velocity of light, and therefore in the index of refraction. The analogy in the two cases is then very close. The light vibrations will occur along two directions only, at right angles to each other, and these will give rise to two waves of greatest and least velocity. No intermediate directions of vibration or velocities occur.

If natural light falls upon a crystal, its vibrations automatically break up into two components along the two directions possible in the crystal. If polarized light falls on the crystal, the same thing in general happens, but the components are no longer equal, and in the cases where the vibrations in the incident polarized light happen to coincide with one or another of the possible directions in the crystal, the wave goes through unaltered.

Uniaxial crystals and the wave surface. Many crystals have one direction, called the *optic axis*, along which a beam of light may be sent without producing any double refraction. If a crystal of calcite is cut with the surface perpendicular to the axis, and a beam of light is sent perpendicularly through it, the crystal acts like a piece of glass. But if a ray is sent slantwise through this plate, double refraction is again observed, the angle of separation between the two rays increasing with the angle of slant.

If a luminous speck should suddenly spring into existence inside a crystal and generate light for a very short time, the advancing wave-front of this disturbance would be found upon a two-sheeted surface called the *wave-surface.* This would consist of a sphere and an ellipsoid of revolution (a figure traced out by revolving an ellipse about one of its axes), and these two would have two common tangent points; that is, one is inscribed in the other. The

sphere may be inside the ellipsoid, as in calcite, or the ellipsoid inside the sphere, as in quartz (Fig. 36–9). The sphere belongs to the ordinary ray; the ellipsoid to the extraordinary. A plane which is tangent to one of these surfaces at any point is a possible wave-front; the line joining the center to the point of tangency is the corresponding ray. In the case of the extraordinary ray, we note the curious fact that the ray is not usually perpendicular to the wave-front.

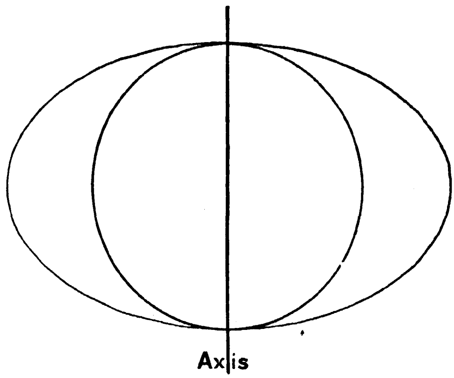


FIG. 36–9
Type of wave-surface for calcite

Sections of the crystal cut in different ways with respect to the axis will act differently, and the directions of the two rays can be worked out by considering sections of the wave-surface drawn to match each case. These details will be omitted. We shall also omit consideration of the more complicated crystals, such as mica, which possess *two axes*, and a wave-surface which is curious and interesting.[1]

The uses of the polariscope. A polariscope (p. 574) usually consists of two Nicol prisms in tandem. When the prisms are crossed, no light passes through the instrument, but in this condition it is particularly useful in the study of crystals. For example,

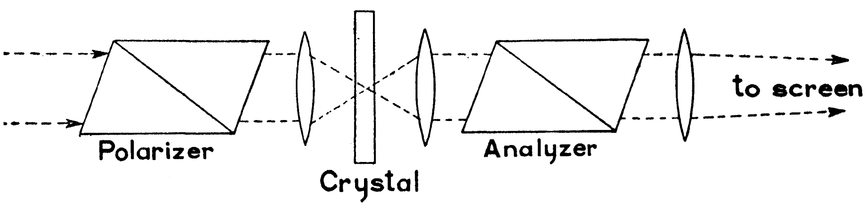


FIG. 36–10
A polariscope for the examination of crystals

quartz occurs in two forms which may be mixed in a single specimen, forming an example of what is called "twinning." Common "pebble" spectacle lenses are ground from quartz crystals and often exhibit this peculiarity, which is quite invisible to the unaided eye. When such a specimen is placed between crossed Nicol prisms (Fig. 36–10), polarized light from the analyzer passes into the quartz, breaks up into two beams traveling with different

[1] On such subjects consult T. Preston, "Theory of Light"; R. W. Wood, "Physical Optics," or E. Edser, "Light for Students" (all published by Macmillan).

speeds, and when it comes out recombines into a new type of vibration, generally elliptically polarized light, which cannot be stopped by the analyzing prism. Thus on the dark field of view of the crossed prisms the specimen appears bright. Moreover, the speeds in the crystal vary with the color, so that some colors may be blocked out and others not, and when the incident light is white, a display of colors over the specimen is usually seen, which may vary gorgeously as one of the prisms or the specimen is turned. Under such circumstances a twinned crystal shows up instantly, as the brightness and color of the different parts vary conspicuously. Figure 36–11 shows such a quartz crystal.

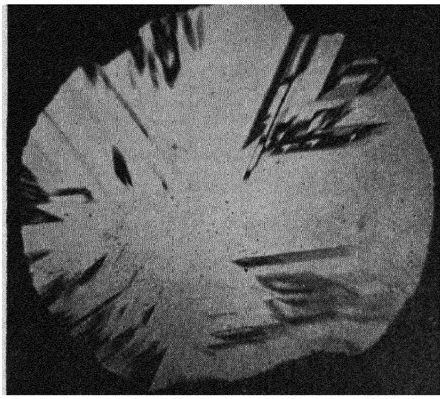


FIG. 36–11
A quartz plate showing twinning, viewed in a polariscope

In order to be able to understand these effects, let us imagine that when the light from the polarizer strikes the crystal, the incident vibration can be represented by AB, Fig. 36–12, while OX and OY are the directions of the possible vibrations in the crystal. The components of AB are then YZ and XW. One of these vibrations gains on the other as they pass through the crystal, and when they recombine on the far side, there may be any difference of phase whatever between them. If the difference is 180°, they combine into ordinary polarized light whose direction of vibration is along RS; but for other angles the result may be any ellipse which can be inscribed in the rectangle $ARBS$, one possible ellipse being shown in the figure. If AB were initially at 45° to OX and the phase difference introduced happened to be 90° (or 270°), the result would be circularly polarized light.

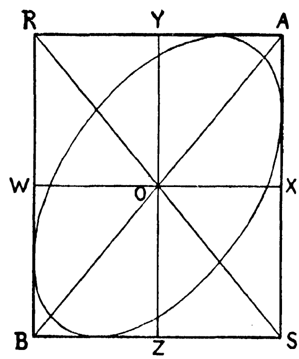


FIG. 36–12

Artificial double refraction. The polariscope also shows most beautifully the existence of mechanical strains in a transparent material. An ordinary paper weight made in the form of a clear cube of glass usually shows a beautiful color pattern near the corners, when it is placed between crossed Nicol prisms. Mechani-

cal strains arise in the material on cooling from the molten condition, which by unsymmetrical pressures impart artificial differences in elasticity to the material in different directions, with which there is always associated a certain amount of double refraction.

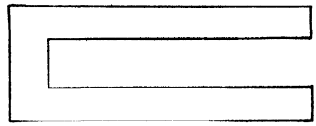


FIG. 36–13

A still more striking case is furnished by a thick (say 5 mm.) piece of transparent celluloid cut out in the form shown in Fig. 36–13. When the free ends are pinched between the fingers while the base is mounted between crossed Nicol prisms, beautiful colored fringes appear whose form is indicated in Fig. 36–14. The study of these indicates the distribution of stresses in the specimen. Small models of structures (e.g., bridges, the framework of dirigibles, gear wheels, etc.) can be made in this way and forces applied to them, carefully chosen to imitate those to which the real structure is to be subjected; and thus experimental predictions can be made as to the behavior of the large-size structures, which may be useful to engineers. This method is most helpful in cases in which the strains are too complicated to be treated mathematically.

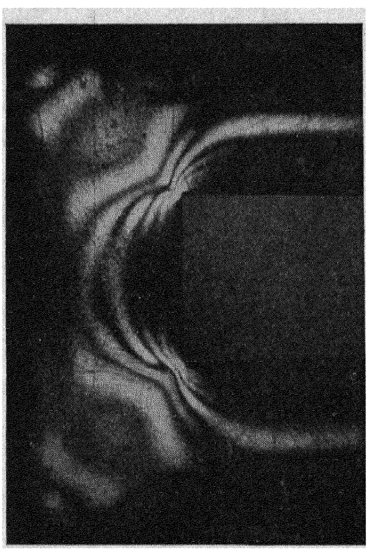


FIG. 36–14
Photograph of strain fringes in a celluloid specimen

Artificial double refraction can also be created in glass by putting it in a strong electrostatic field. This is probably due to the mechanical strain which is caused by the field (p. 319).

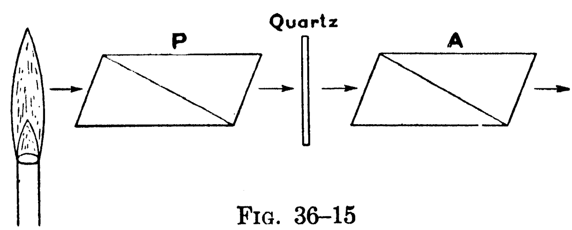


FIG. 36–15

Rotation of the direction of vibration. When a polariscope is set up, as in Fig. 36–15, illuminated by sodium light, with the Nicol prisms carefully set in the crossed position, so that no light gets through, and a plate of quartz cut

perpendicular to its axis is then set between the Nicol prisms, it is found that the analyzer has to be turned through a certain angle in order to blot out the light again. In other words, the quartz crystal has turned the direction of vibration of the light passing through it. A solution of sugar will do the same sort of thing, and the amount of turning will give an indication of the strength of the solution. These measurements are actually used commercially, for instance by customs officials to determine the duty on syrups, but the direction of turning is to the right with one sort of sugar, and to the left with another; so that the matter is not so simple as it might be. Polariscopes for such measurements are called *saccharimeters*.

When the experiment is performed with white light instead of the sodium flame, it is found that color effects are produced, which make exact measurements difficult.

Effects of magnetism on light. The experiment just mentioned resembles a remarkable discovery made by Faraday, which shows a connection between a magnetic field and the direction of vibration in a light wave. In this case a

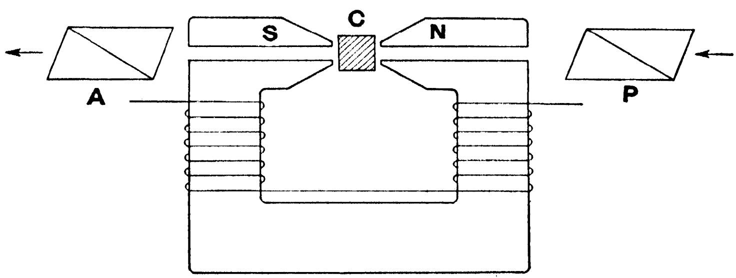


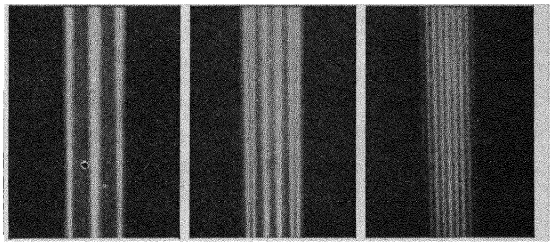
FIG. 36–16

small cube C of dense glass is placed between the poles of an electromagnet and the poles are bored out so that it is possible to look through them in the direction of the magnetic lines, as in Fig. 36–16. A Nicol prism P is placed in front of the magnet and another beyond it as analyzer, A. If the Nicol prisms are crossed before the field is turned on, the presence of the field in the dense glass turns the direction of vibration of the polarized light passing through it through an angle which depends on the strength of the field and the frequency of the light vibration. This *Faraday effect* differs from the rotation produced in quartz or sugar; if a mirror is so placed as to return the beam through the specimen again, the Faraday effect is doubled, while the rotation in quartz is undone. It is as though the quartz rotation were connected with a corkscrew action of some sort which unscrewed itself when reversed; while the magnetic field produced a rotation in space, connected perhaps with the rotation of the current in the magnetizing coil, which acts in the same sense whichever way the light is going.

The Zeeman effect is a change in the lines observed in the spectrum of a source of light when the source itself is placed in a strong magnetic field. This effect also was looked for by Faraday, but it was not found till 1896 when Zeeman[1]

[1] P. Zeeman, professor of physics at the University of Amsterdam.

discovered it with better apparatus. If a short spark is formed in a very strong magnetic field and the light is observed in a direction at right angles to the magnetic lines, a single spectrum line is seen to be broken up into a close group as in Fig. 36-17, which may take the form of a triplet (*a*) or something more complicated (*b* and *c*). The components of the group are polarized, the outer ones vibrating perpendicularly to the magnetic lines, the inner ones parallel. The theory of this effect, based on the behavior of atomic electrons in the field, has led to results in harmony with the electron theory of matter.



(*a*) (*b*) (*c*)

FIG. 36-17

Three different lines in the zinc spectrum which are ordinarily single, but break up into these patterns when the source (a spark) is in a strong magnetic field.

The existence of a general magnetic field on the sun has been demonstrated by means of the Zeeman effect, and very large fields have been shown in the same way to be created by the huge storms which we see as sun-spots.

CHAPTER 37

X-RAYS AND CRYSTAL STRUCTURE [1]

The discovery of X-rays. It is curious to think that energetic and ingenious experimenters could have been busily investigating the phenomena of nature for centuries and yet have remained entirely ignorant of the existence of a whole class of rays of the utmost practical and scientific importance. The actual discovery was finally made almost by accident and created a tremendous sensation in the scientific world. In 1895 when Roentgen [2] was studying the discharge of electricity through gases, he noticed that a certain tube covered with black paper caused a screen coated with a fluorescent material (barium platino-cyanide) to glow brightly in the dark, even at a distance of several feet. A new sort of rays was evidently issuing from the tube and passing through the black paper, which he named X-rays to indicate that their origin was unknown. He soon found that they came from that part of the glass wall of the tube where cathode rays (p. 441) were being stopped. When the cathode rays were concentrated on a small spot inside the tube and a special metal target was placed there to stop the rays, this point served as a source of the rays, just as though it were a narrow source of light, casting sharp shadows of objects placed in front of the fluorescent screen. This showed that the X-rays travel in straight lines from such a source. Roentgen also found that they are more absorbed by dense matter than by that which is lighter, and that they affect photographic

[1] This chapter and those following may be omitted in a short course, or by those who wish to dodge the complex but interesting issues raised by modern physics.

[2] Wilhelm K. Roentgen (1845–1923), then professor of physics at the University at Würzburg.

plates, so that permanent records can easily be made by means of them. X-rays are differently absorbed by flesh and bones; thus shadow pictures ("radiographs") can be made of the bones of the hand, leg, etc., which have been of great assistance in the

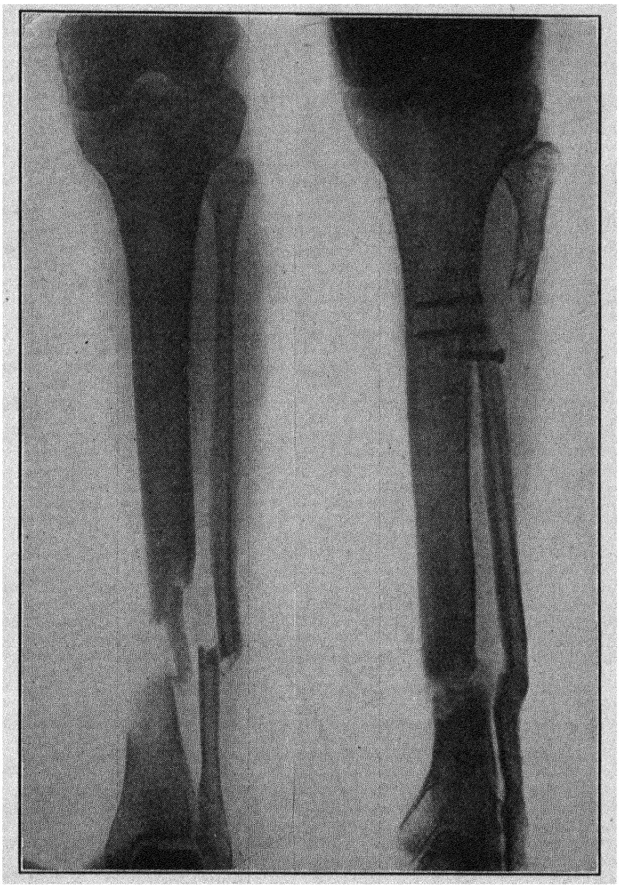


FIG. 37–1

An accident broke both leg bones (left), and, as the main one failed to heal properly, the surgeon purposely broke the other near the top, and (right) screwed it to the first, in order to furnish the necessary stiffness. (Courtesy of the Massachusetts General Hospital)

treatment of human ills (see Fig. 37–1). On the other hand, the first experiments indicated that X-rays are unaffected by lenses, mirrors, prisms, magnetic fields, or any of the other agencies to which one might naturally try exposing them. To explain such independence of behavior, several theories as to the nature of X-rays were proposed, but the true explanation was not found for

many years, when it was shown that they are electromagnetic waves of the same nature as light waves, but from one to ten thousand times shorter in wave-length. We shall next consider these fundamental experiments.

The diffraction of X-rays. As some evidence pointing to the conclusion that X-rays are really very short light waves had already been obtained by experiments on diffraction through narrow slits, it occurred to Laue [1] in 1913 that a diffraction grating

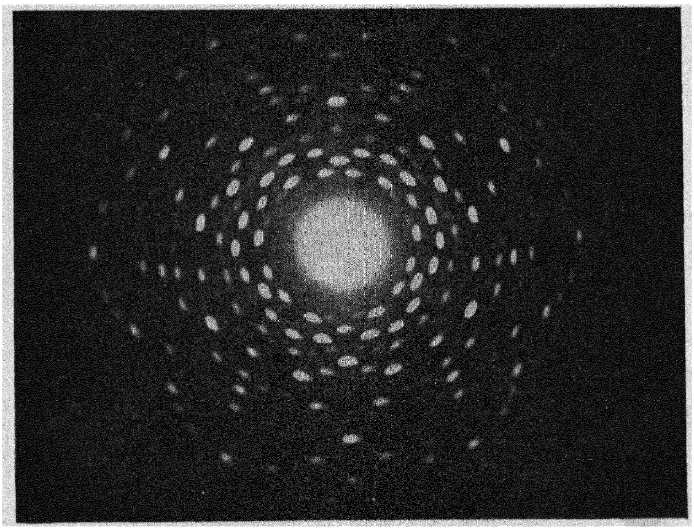


FIG. 37–2
A Laue pattern for quartz. (Courtesy of Dr. R. W. G. Wyckoff of the Rockefeller Institute for Medical Research)

ought to act on them in the usual way if it were only made on a small enough scale, and that a crystal, being an arrangement of atoms in orderly fashion, might be just such a grating already made for such use. Accordingly a narrow beam of X-rays was sent through a crystal of zinc sulphide to a photographic plate on which there was found a record of the direct beam, as a central spot, surrounded by a series of fainter spots distributed in a strikingly symmetrical pattern which varied with the sort of crystal used. Figure 37–2 shows such a pattern obtained with a plate of quartz. The pattern is different with other materials. This ex-

[1] Max von Laue, then at the University at Munich; now professor of theoretical physics at Berlin. The actual experiment was carried out by Friedrich and Knipping.

periment opened up a new experimental technique from which the wave-lengths of the X-rays could be measured. The spots are produced by diffraction, but the explanation is not simple. We shall consider instead a more direct method.

Measurement of X-ray wave-lengths. Crystal diffraction. A crystal is an orderly arrangement of particles in space, in a sort of three-dimensional lattice. Such a "space-lattice" has planes in which the particles are more abundant than elsewhere, and these planes may be inclined at various angles. Figure 37-3 shows a regular arrangement of particles, and one set of planes that can be drawn through them. If A and B are two rays from the incident beam, which are partially reflected at P and Q, it is evident that the wave sent off at the angle of reflection from Q will be behind that from P by a distance $LQM = 2LQ = 2PQ \sin LPQ = 2d \sin \theta$, where d is the distance between the planes and θ is the angle with the surface, sometimes called the "glancing angle." If these two waves are in the same phase, they will form part of a wave-front going off at this angle, and the waves from all other planes (R, S, etc.) will also be in phase. Hence a fairly strong reflected wave will be found, as in the case of the grating (p. 552), if the condition necessary for equality of phase is satisfied; namely that the retardation from one ray to the next should be a whole number of wave-lengths, $n\lambda$. Hence we reach Bragg's[1] equation

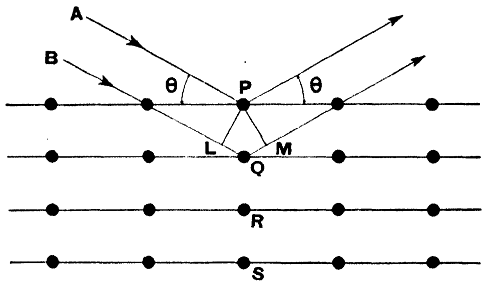


FIG. 37-3

$$2d \sin \theta = n\lambda$$

for the directions in which the X-ray beams will be found which are diffracted by such a three-dimensional grating.

As a result of this equation, the wave-length of X-rays can be measured directly from an observed angle and the "grating space" d. The determination of the latter can be made from the actual

[1] Sir William Bragg, professor first in Australia, later in Leeds and London. With his son, Professor W. L. Bragg of Manchester, he originated this beautiful method of study of the nature both of X-rays and of crystals.

weights of the atoms involved and the density of the substance.[1] In this way Bragg found, for instance, that the distance between the planes parallel to one of the cubic faces in rock salt is 2.81 $\times 10^{-8}$ cm. or 2.81 A. U. The results of the wave-length measurements yield values for the easily absorbed ("soft") X-rays of 5 to 10 A. U., while penetrating ("hard") X-rays are shorter, running down to about 0.1 A. U. It will be recalled that the wave-lengths of visible light lie in the range 7500 to 4000 A. U. As diffraction phenomena shrink to a smaller scale when the wave-lengths are shortened, we see that X-ray diffraction effects are on a scale thousands of times smaller than those of light. This explains why they remained so long undiscovered. On account of their shortness, a new unit of measurement is coming into use. An X-ray wave-length of 1 A. U. is written as 1000 "X. U."

If the wave-length of the X-rays is definite and known, the crystal may be turned about at all angles until reflection is obtained from the planes in which its atoms are concentrated, and thus the separations of these planes can be found. The intensities of the reflected beams depend on the numbers of particles contributing energy to them. As the distances between the planes of atoms are thus found and also the relative numbers of atoms in these planes, it has been possible to find the arrangement of the atoms in the space-lattice of all common crystals, including crystals of pure metals, as well as alloys and compounds. These interesting results lead one into the subject of crystallography, where we shall not have time to pursue them.[2]

[1] In rock salt, for instance, one atom each of sodium and chlorine are found in each pair of unit cubes out of which the crystal is made. There are 6×10^{23} molecules (p. 146) in a "gram-molecule" of any substance; or in this case in 58.46 grams (the sum of the atomic weights of sodium and chlorine gives this number). The density 2.18 is the mass of two unit cubes divided by its volume ($2L^3$). Hence $2.18 = \dfrac{58.46}{6 \times 10^{23} \times 2L^3}$; from which $L = 2.8 \times 10^{-8}$ cm., approximately.

[2] Those who desire to see how simply and intelligibly a piece of research *may* be presented by its author should read the popular book "Concerning the Nature of Things" by Sir William Bragg, 1925 (Harper) and the somewhat more complete account in "X-rays and Crystal Structure," by W. H. and W. L. Bragg, 1923 (Bell) and "An Introduction to Crystal Analysis," by W. H. Bragg, 1929 (Van Nostrand). See also "X-rays" by G. W. C. Kaye, 1923, (Longmans, Green) and "Applied X-rays" by G. L. Clark, 1927 (McGraw-Hill).

Ionization of gases by X-rays. Roentgen observed that bodies lost their charge in the neighborhood of an active X-ray tube, and this has been traced to a disrupting action on the part of the rays when they are absorbed by the molecules of air. Their energy, like that of ordinary light, appears to be emitted and absorbed in quanta (p. 545) and when an atom absorbs a large quantum, such as X-rays furnish, what happens is that one of the *inner* electrons is ejected from the atom. Outer electrons then fall in to take the place of the missing one, and during this rearrangement other X-rays are emitted by the atom. Finally the capture of an outermost electron restores the atom to its normal neutral condition. While it exists as an ion, however, it does its share in making the gas of which it is a part a conductor of electricity.

A charged electroscope may be placed across a large room from a strong source of X-rays, and when the rays are generated the leaf of the electroscope begins to fall at once, indicating the formation of many ions by absorption of the X-rays in the air near it. When the X-ray tube is shut off, the leaf slows down and ceases its motion. It does not stop instantly, as it takes an appreciable time for the last of the ions to disappear by recombination. The rate of fall of the leaf may be used as a quantitative measure of the intensity of the X-rays.

Modern X-ray tubes. Since 1913 the form of X-ray tube invented by Coolidge [1] has come into universal use. In the earlier forms of tube the stream of electrons depended mainly on the ionization of the small amount of gas left in the tube. In the Coolidge form (Fig. 37–4) the tube is highly evacuated and the electrons come from a heated filament F as thermions (p. 451). Their number can be varied by changing the heating current sent through this filament. Their speed, however, depends on the high difference of potential separately maintained between F and A, the metallic target on which the electrons fall. A high speed on the part of the electrons involves more loss of kinetic energy by impact with the atoms of the target, larger quanta of energy, and therefore higher frequency of the waves sent out by them. Thus the penetrating power of the X-rays is controlled. The pressure in the tube is about a millionth of a millimeter of mercury.

[1] Dr. W. D. Coolidge, Assistant Director of the General Electric Research Laboratory at Schenectady, N. Y.

Practical uses of X-rays. The surgeon finds X-rays a convenience when he deals with broken bones. The modern dentist could not think of getting along without their help. Penetrating rays disclose tubercular defects in the lungs. After a man has swallowed a large dose of a heavy (and X-ray absorbent) but harmless salt, such as barium sulphate, his entire digestive tract becomes easily observable, and irregularities in its behavior are evident. Motion pictures of the internal organs are a recent development.

In all X-ray work the operator must be protected from the rays by sheets of lead or of lead glass. Before the necessity of these precautions was realized, dangerous external injuries somewhat like burns, and internal ills of an even more serious nature led to

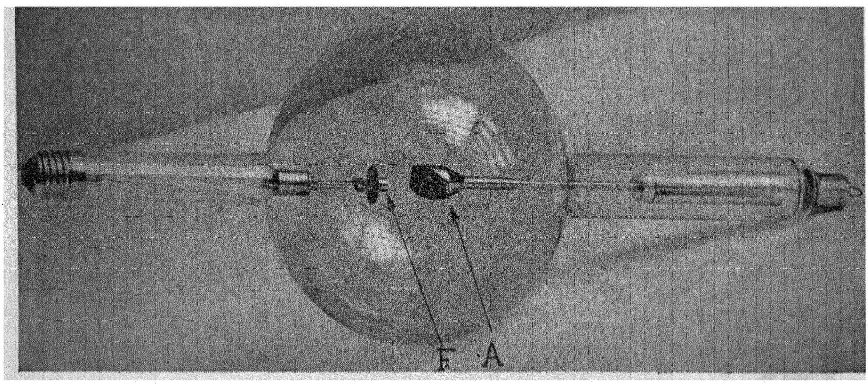


FIG. 37-4
A Coolidge X-ray tube. (Courtesy of the General Electric Company)

much suffering, and even death, among early experimenters with X-rays. The destructive action which penetrating X-rays have upon certain kinds of tissue is now utilized in treatments which destroy or check malignant growths within the body. Biologists, too, have tried X-rays on insects and found that they produce curious effects in the cells which determine hereditary characteristics, thus actually causing variations of form in offspring, called "mutations."

In the manufacture of machinery, X-rays may disclose blowholes in castings or internal faults in important steel bars, such as the shafts of airplane engines. In art, too, these rays are now of great use in detecting changes in old oil paintings, the original often showing through in spite of a more modern covering.

Many difficulties present themselves in X-ray photography.

For instance, if the inside of a thick bar of steel is to be studied, very penetrating rays must be used, and these will pass through a photographic plate without leaving much effect behind; in other words, the photograph will be faint. This trouble is overcome by the use of a uniform plate made of a fine-grained, heavy fluorescent salt, such as calcium tungstate, which is placed in close contact with the photographic film. The X-rays are absorbed by the salt and make it glow, and this light affects the plate; thus the picture may be greatly "intensified."

X-ray spectra. When the spectrum given by an X-ray tube is to be examined, this is done by sending the rays from a narrow source through a slit to a crystal, whose spacing is known, and then measuring the "glancing angles" at which the X-rays are reflected. These are detected photographically or by means of an "ionization chamber," (Fig. 37–5) which is a small metal vessel in which the rays produce ions and thus cause a current to flow in the measuring circuit.

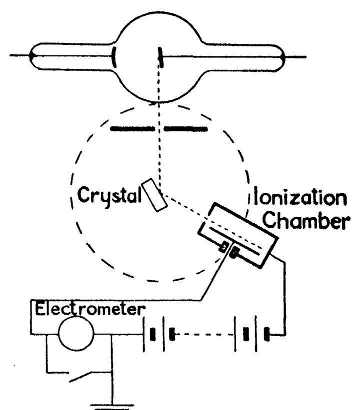


FIG. 37–5

Thus it is found that X-rays give both ***continuous*** and ***"bright-line"*** spectra, as ordinary light waves do; ***absorption*** spectra, also, are produced by many substances. The continuous spectra have a definite upper limit, whose existence is interesting. The high voltage put on the tube has, of course, a maximum value. The energy an electron acquires in falling through this difference of potential may be emitted as a single quantum, and being a large quantum, it will involve a high frequency, or short wave-length. But evidently, if the quantum theory is true, there will be a sharp limit to the spectrum on the short-wave side beyond which no X-rays are emitted (see curves in Fig. 37–6); since if any were there they would involve more energy than the tube could furnish. This particular fact has been carefully tested. The high-frequency limit has been found to be strictly proportional to the voltage put on the tube, just as it should be according to this theory.

Figure 37–6 gives curves showing the intensity at each wavelength of the continuous spectrum of an X-ray tube (with a target

of platinum and nickel) at three different voltages. Rising from the uniform curve of the continuous spectrum one notes peaks which are due to "bright-line" emission spectra. These prove to be characteristic of the metals composing the target where the impact of the cathode rays in the tube generates the X-rays. Such peaks have been found to appear in groups, the heavier elements giving groups called the *K* series, the *L* series, the *M* series, etc. (The term "series" as used here has no connection with the line series of p. 541.) The lighter elements give fewer of these series. Of these the *K* series is the simplest group. It is now known to be emitted when an electron has been dislodged from the innermost part of the atom and is being replaced by one falling in from the next outer "shell" to take its place. In fact the innermost shell of electrons in the atom is sometimes called the *K* shell, the next the *L* shell, etc. The description given on p. 587 of the ionization of an atom by X-rays can now be reworded thus: a high-speed collision ejects a *K* electron from the atom; an *L* electron falls in to take its place, causing the emission of a *K* line in the spectrum; an *M* electron then falls in to the vacancy in the *L* shell, emitting an *L* line in the process, and so on until the vacancy occurs among the outermost electrons of the atom. It then picks up an electron from outside, and settles down in its normal state, complete once more. Evidently if this theory is correct, a very light atom, with electrons in two shells only, could emit nothing but *K* series X-rays; one with three shells, *K* and *L* series; and so on, and this is actually found to be true.

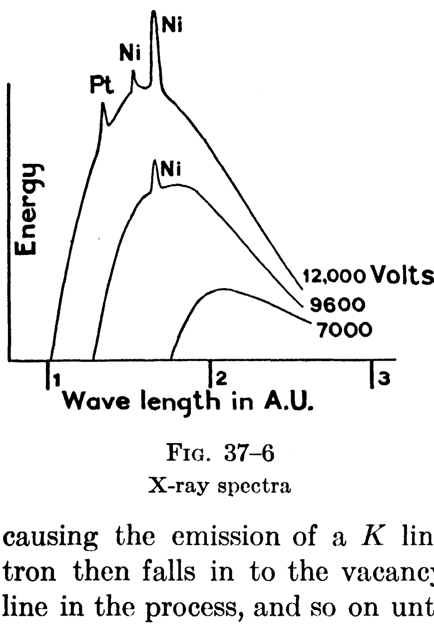


FIG. 37–6
X-ray spectra

Moseley's law. Atomic numbers. The strongest evidence in favor of the picture just given of the arrangement of electrons in an atom comes from the researches of Moseley [1] in 1913 on the wave-

[1] H. G. J. Moseley; studied under Rutherford in the University of Manchester; did research work there and in Oxford, and was killed in the war in

lengths in the X-ray spectra of many elements. He found that the K series spectra were alike in their arrangement in all the spectra tested, but their position on the frequency scale varied with the position of the element in the chemical table, as indicated in Fig. 37–7. It had already been suggested from other experiments that each atom contains a number of electrons approximately equal to half its atomic weight. This implies, for instance, that nitrogen (atomic weight, 14) has seven electrons, oxygen (atomic weight, 16) eight, silicon (atomic weight, 28) fourteen, and so on. If the

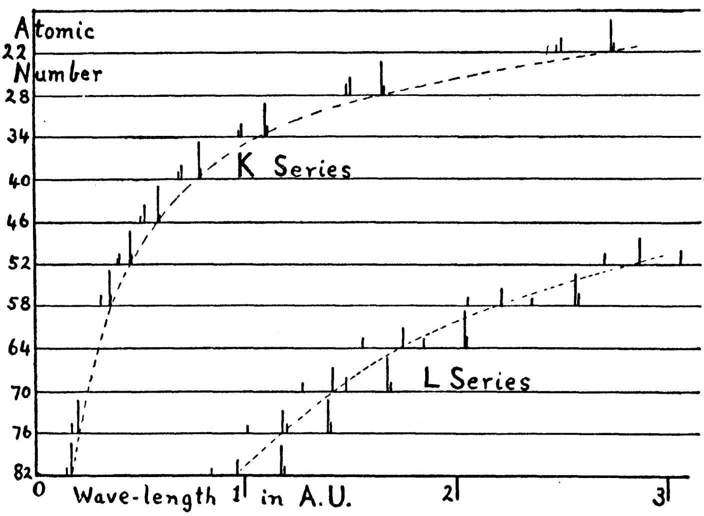


FIG. 37–7
K series X-ray spectra of several elements

chemical elements are arranged in the order of their atomic weghts, their position in the table is expressed by this same number; thus nitrogen is the seventh element, and silicon the fourteenth. Moseley established the term ***atomic number*** for the number of electrons belonging to each kind of atom, or the number of the element in the chemical table. He did this by discovering that curves like those in Fig. 37–8 could be drawn expressing the law of arrangement of the X-ray spectra among different elements, in which the ordinates are the square roots of the frequencies of one of the K (or L) series lines in all the elements in a certain range of atomic number, and the abscissæ are the atomic numbers themselves.

1915, having finished at the age of twenty-seven such a mass of important work as to ensure him a lasting fame.

The curves are almost straight lines, and the points fit perfectly. If atomic weight is used instead of atomic number, the points do not fit so well; there are, in fact, some definite misfits; this reveals the fact that the atomic numbers are more fundamental characteristics of the elements than the atomic weights.

This discovery makes us quite sure as to the number of chemical elements that exist. There were gaps in our knowledge in Moseley's time; for instance, the element of atomic number, 72 (hafnium) had not yet been discovered, but his work made its X-ray spectrum known in advance. When a small amount of this material was finally obtained, the predictions based on his law were completely verified.

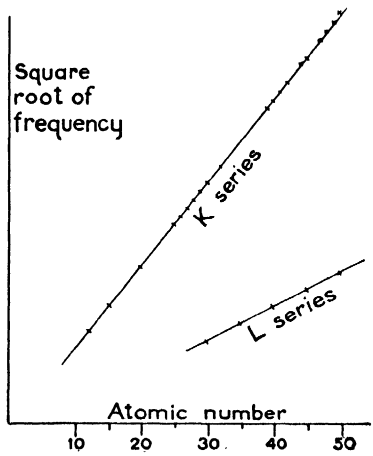


FIG. 37–8

Atomic arrangements of electrons. Moseley's work established the fact that the innermost parts of all atoms (heavier than helium) are alike. There are just two electrons in the *K* "shell," the same two which make the helium atom a structure so strong and self-contained that it is chemically inert; that is, it has no appreciable external electric field with which to act on other bodies. When the results of the study of optical spectra are combined with those on X-rays and with the known chemical properties of atoms, we are able to say that the *L* shell can contain as many as eight electrons, but no more, as this number produces another very stable atom, that of the inert gas ***neon.*** When eight more are put in the *M* shell, we arrive at a third inert gas, ***argon.*** Then eighteen must be added before we complete the *N* shell with the inert gas ***krypton;*** and so on.[1]

The Compton effect. In recent years (1923) A. H. Compton [2] has carried out experiments which give extraordinary support to the quantum theory. He

[1] The reader who is familiar with chemistry will find in Richtmyer's "Introduction to Modern Physics" (Chapter 11) 1928 (McGraw-Hill) a very readable account of this fascinating part of modern physics.

[2] Professor of physics at the University of Chicago; winner of the Nobel prize for physics in 1927; author of an excellent book on "X-rays and Electrons," 1926 (Van Nostrand), to which the reader may turn for a full account of X-ray research.

found that when an X-ray beam, consisting of a single wave-length, strikes a piece of some material, carbon, for example, there are *two* wave-lengths scattered from the carbon, one of which is the same as that of the incident ray, while the other is new. The difference in wave-length depends on the angle of scattering. To put it in a greatly exaggerated way, the effect is like an imaginary experiment in which blue light falling on a scattering medium sends off at one angle a mixture of the original blue *plus green;* at another angle, blue *plus yellow;* and at a third, blue *plus red.* Of course, such large changes of wave-length as these do not actually occur, but the real effect is no more unlike an ordinary mechanical process than the one we have just imagined. No explanation based on familiar wave-theory terms can be suggested. On the other hand, the quantum theory, applied to the scattering collisions with the help of the principle of the conservation of energy and of momentum, furnishes a very simple explanation. The incident quantum of X-ray radiation is to be treated like an ordinary object, or bullet, which is capable of colliding with something else and rebounding. It contains a certain amount of energy, E, appropriate to its frequency, f, according to the equation $E = hf$. If it strikes an atom, and an electron absorbs some of this energy in the "collision," the quantum of X-rays, which thereby rebounds (or is scattered) from the electron, must inevitably have less energy than before, and therefore a lower frequency, or longer wave-length. Also, the electron receiving the impact must recoil under it; the atom is ionized, and the electron is given energy enough to escape at a speed which varies with the angle. These "recoil electrons" have actually been observed. In this explanation we must boldly treat the quantum as though it resembled a bouncing tennis ball, while at the same time it remains a train of waves. To those who wish nature to show herself continually as more and more simple and easy to understand, these results are no comfort; and matters are only made worse by two other discoveries which we must consider next.

The Raman effect. An effect somewhat similar to that found by Compton was discovered in 1928 in the ordinary optical part of the spectrum by Raman.[1] A monochromatic beam of light (i.e., one containing a single wave-length only) when concentrated in great strength on any one of a variety of substances (a quartz crystal, liquid benzol, or even a gas such as hydrochloric acid) will produce a faint scattered light (so faint, in fact, that special precautions may have to be taken before it can be observed) in which is found the original wave-length, and also certain new (and usually longer) wave-lengths, which are sharp and definite. These are caused, after the manner of the Compton effect, by collisions between quanta of radiation and molecules of matter. Here the incident quanta are much smaller than in the case of the Compton effect, and they cannot remove electrons from the atoms, but only excite each atom into one of its stationary states. The scattering material has certain natural absorption frequencies and can therefore absorb quanta of a definite size, and these only. If one of these is absorbed in a collision, its amount of energy is subtracted from that in the incident quantum, leaving a diminished quantum with which to produce the scattered light, which must then consist of a spectrum line of longer wave-length. In rare cases a molecule of the scat-

[1] C. V. Raman, professor of physics in the University of Calcutta.

tering substance might have been "excited" in this way by an earlier collision, and instead of losing its energy as a quantum of radiation of long wave-length, it might keep it a while and then give it by collision to a new incident quantum so as to send this off with *more* than it had before. This would account for the very faint lines observed in some cases on the high-frequency side of the main exciting line.

If the active molecules were ionized, as the atoms are in the Compton effect, the ejected electron might carry off with it any amount of energy, depending on its speed, and so absorb from the incident quantum an amount capable of gradual change with the angle. But in the Raman effect this does not occur; the molecule is merely excited to a higher stationary state, and thus definite energy changes and sharp "Raman lines" are the result.

The diffraction of electrons. Another extraordinary experiment, done in 1927, must now be noted, as it furnishes a direct connection between electrons and wave-motion, somewhat akin to the effects just considered. This was the diffraction of electrons by a crystal, and was first carried out by Davisson and Germer.[1] In one of these experiments a narrow stream of electrons was emitted from a hot filament, passed through an accelerating electric field which could be varied, and fell upon a single crystal of nickel. The number of electrons reflected, or scattered, at various angles was measured by collecting them at each angle and measuring the current they produced. The results can be best described by saying that the electrons behaved exactly as though they consisted of very short waves and were diffracted by the crystal, acting like an ordinary diffraction grating, through angles which depended on the ***speed*** of the electrons. Their wave-lengths were obtained from the angles and the "grating space" of the crystal, and proved to be of the order of ordinary X-rays (about 1 A. U.), but their wave-length varied with their speed. Thus it appears either that electrons ***are*** waves, which seems impossible in view of their other properties, or that ***they carry waves with them,*** and go wherever the waves travel. If electrons act like waves, and radiation quanta act like particles, we are faced with a very novel situation, and can only hope that a genius may soon arise to show us how to harmonize these apparently contradictory properties. Although we cannot imagine such a situation, we should be careful not to follow the example of Descartes, who argued that a vacuum is impossible because it is inconceivable.

[1] Drs. C. J. Davisson and L. H. Germer of the research division of the Bell Telephone Laboratories in New York City.

Wave mechanics. A new theory has recently arisen which deals with electrons and atoms in terms of the waves just considered. The theory is couched in rather general mathematical terms and does not furnish an atomic model that can be visualized, but it has been very successful and stimulating. It was started by L. de Broglie in Paris, developed by Schrödinger (now of Berlin), and is now growing actively in the hands of a host of younger workers.

The waves in the experiment of Davisson and Germer determine the abundance of electrons at each angle. They may also be thought of as waves of probability, since their intensity in any region indicates the probability that electrons will be found in that region. The new theory states that the distribution of electrons anywhere, and in particular within an atom, depends upon a system of such probability waves. These may have a high intensity inside an atom and practically none outside, just as the set of waves created by a stone dropping into water may have a localized intensity at any moment, none being ahead of the close group of growing rings, and none behind, where the disturbance has subsided. The theory does not encourage us in attempting to fix the position of one electron at any time. Heisenberg's "principle of uncertainty" indicates that we can never precisely know both the position and the speed of an electron at the same time. This comes from the fact that we cannot examine an electron by any method that does not involve disturbing it by some physical agency which has a definite "grain," or degree of coarseness. If we cannot ever know these minute details about single electrons, perhaps we must content ourselves with statistical information in regard to hordes of them.

The theory does not yield the distribution of electrons about the nucleus of an atom as a definite picture, as in Bohr's theory. Instead, it shows how the probability of the occurrence of electrons in a given point within the atom varies with its distance from the center of the atom, all of which is determined from the properties of groups of waves. Inside an atom these form into standing-wave patterns which take certain shapes only. Just as one can divide a vibrating string into a few sets of standing waves (p. 236), the intermediate cases being impossible, so inside an atom certain distributions of electrons are possible, governed by the standing-wave patterns, and no others can occur. These then must be the stationary states of the atom, here reached with less magic, though perhaps less concreteness, than in Bohr's theory. Happily, the new theory is able to show that in the case of hydrogen, the amounts of energy in each of these atomic stationary states is precisely the amount given by Bohr, and required for the explanation of atomic spectra. In many other directions also the wave-mechanics theory has been conspicuously successful. While theories of the atom have not in the past lived very long, this one is still in a state of very healthy growth, which will doubtless continue until the blazing light of a still better one causes it to wilt and disappear.

CHAPTER 38

PHOTOELECTRICITY AND LUMINESCENCE[1]

The discovery of the photoelectric effect, 596; the nature of the effect, 596; source of the ejected electrons, 597; the velocities of the photoelectrons, 597; photoelectric cells, 599; practical applications, 600; photoelectric theory of photographic action, 602; luminescence, 603; explanation of fluorescence and phosphorescence, 603.

The discovery of the photoelectric effect. In connection with work on the reception of electromagnetic waves in 1888, Hertz noticed that sparks in air between two pieces of zinc would jump a longer distance when the spark terminals were illuminated by light than when they were in the dark. He traced the difference to the presence of ultra-violet light, found that this light must fall on the metal of the terminals and showed that the effect was greatest if the surface of the metal was brightly polished. It was later found that a negative charge on an insulated plate of zinc disappeared quickly when ultra-violet light fell upon the plate, while a positive charge did not, an experiment which is easily repeated in a classroom with an arc lamp and an electrometer. This is now known as the ***photoelectric effect*** and is found to be produced not only by ultra-violet light but by X-rays, and under proper conditions by visible light also. Since the time of Hertz, a large amount of research has been done on this subject, with results so remarkable both in their effect on atomic theory and in their practical applications that we must consider them in some detail.

The nature of the effect. One who is familiar with the electron theory can readily explain these early experiments by assuming that the photoelectric effect is due to the escape of electrons from the zinc surface when ultra-violet light falls upon it. These electrons ionize the air in the experiment of Hertz and make the passage of the spark easy. They will escape from the metal when it is charged negatively, or when it is neutral, but when it bears a considerable positive charge, they are no longer able to get away against the resulting attraction. In a vacuum, these escaping

[1] This chapter may be omitted in a short course.

electrons can be bent like cathode rays (p. 441) by a magnetic field, and their velocity and the ratio of charge to mass (e/m) can be measured. In this way the conjecture that they are electrons was abundantly verified, and they were found to be moving at about the same speeds as the particles in cathode rays, these speeds varying curiously with the conditions of the experiment.

Source of the ejected electrons. A question immediately arises in regard to these electrons. Do they come from within the atoms, or are they the "free" electrons which are supposed to be responsible for electric conduction? Two lines of argument lead us to prefer the first of these alternatives.

The free electrons presumably share the temperature agitation of the particles of the metal when the latter is heated. Certainly they can escape as thermions (p. 451) when the temperature becomes high enough. Hence if they were the ones ejected by ultraviolet light, they should come off more abundantly from a hot surface than from a cold one. While experiments on this subject are difficult, we now know that this is not so.

Furthermore, those metals which are the best conductors of electricity (such as copper and silver) and therefore have the largest numbers of free electrons, are ***not*** the ones that show the greatest photoelectric action. On the contrary, the action seems to be connected with the peculiarities of atomic or molecular structure. Those metals are most active which have few electrons in their outermost "shell," notably those that have only one, such as sodium, potassium, etc. Magnesium, zinc, and mercury, having two outer electrons, are also active. So are other metals, slightly, and certain compounds, notably potassium hydride, to a great extent.

As we shall see we must conclude that the energy of the photoelectrons comes from the light beam and not from the atom itself, and that this energy is emitted and absorbed in ***quanta.***

The velocities of the photoelectrons. The picture of the photoelectric process to which the above ideas lead is that the beam of light strikes the atoms in the surface of the metal, penetrating through a thin layer, perhaps five or ten atoms deep. A very small number of these atoms absorb a quantum each from it, and this energy is sufficient in each atom to eject an electron at a very high speed. In any one case this speed is the same for all atoms, but those electrons which have to work their way out from lower layers may emerge with diminished speed, having lost energy

through collisions on the way. Owing to the huge number of atoms in all, only a very minute fraction of the atoms need be active at any one time in order to supply the observed current. If one atom out of every hundred million ejects an electron in a second, this will form a strong photoelectric emission. It would then take an impossibly long time for a plate of metal to be deprived of any considerable proportion of its electrons, and there is no definite evidence that the electrons in a piece of metal can be "used up" in this way.

Examining the photoelectric effect with preconceived notions derived from ordinary mechanics, one would be inclined to say that all the particles ought to be affected alike by an incident light wave, that those of the same natural period could pick up energy by resonance and so in time acquire a large amount, perhaps enough to cause them to fly apart; and, finally, one would be sure that the amplitude of oscillation (and therefore presumably the speed of the ejected electron) must depend on the intensity of the incident light. ***These conclusions are all wrong.*** It has been shown that a very small percentage of the atoms is affected at all, that the ejection can occur in a millionth of a second after the light is turned on (which is not long enough for it to be caused by resonance) and that the number of the ejected electrons, and ***not*** their velocity, varies with the light intensity. This last fact makes the photoelectric "cell" (p. 599) a useful quantitative device, of which many applications have already been made.

The most curious fact, however, is that the speed of the ejected electrons depends on the ***frequency*** of the incident light. A faint beam of ultra-violet light sends off a few high-speed electrons; a strong beam of visible light may eject many low-speed electrons, but we cannot speed them up by illuminating the surface more strongly. We now imagine that the radiation travels in quanta, and each quantum has an amount of energy which depends on its frequency. We formerly (p. 545) expressed this by the equation $E = hf$, but now we must admit that the electron escaping from a solid may lose a small quantity of energy, W, as work done in escaping from the surface. Hence the kinetic energy of an ejected electron is

$$\tfrac{1}{2}mv^2 = hf - W,$$

where m is its mass and h is Planck's constant. This equation was proposed by Einstein in 1905.

To show the essentially unmechanical nature of this law we might use an analogy given by Bragg (here somewhat modified). Let us suppose that a series of waves is started on the surface of still water on which a large number of corks are floating. If the experiment followed quantum laws, where the waves were vigorous many corks would jump, at a distance only a few would do so; nevertheless, ***all corks that jumped at all would have the same energy and therefore jump to the same height.*** This height must depend not at all on the amplitude of the waves but only on their frequency. Also, many corks both far and near must keep perfectly still all the time (because less than a whole quantum cannot be absorbed by an atom). There seems hardly room in this picture for the waves which are supposed to initiate the whole disturbance. In fact, it might be better to suppose instead that the energy is carried by bullets which knock the corks up into the air when they hit them, but otherwise leave them untouched. If the speed of such bullets were constant, they should affect the corks alike at all distances. But now there seems no way of arranging the bullets and the corks so as to bring in the observed change of velocity with frequency. In fact the idea of frequency has no association with bullets at all. Thus we have still no working model of the photoelectric effect.

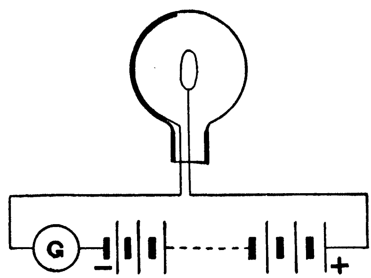


FIG. 38–1

Photoelectric cells. A photoelectric cell is a partially or completely evacuated glass vessel, on the wall of which is deposited a sensitive metal film, connected through an outer terminal to a galvanometer, and containing a grid, or ring, through which the light falls upon the sensitive surface, and to which the liberated electrons are attracted. A difference of potential is maintained between the ring and the metal surface, the ring being positive, as shown in Fig. 38–1. The galvanometer is connected in the same circuit with the battery which furnishes this E. M. F., and measures the current carried by the electrons. The form of cell used in television apparatus is shown in Fig. 38–2. Various materials are used for sensitive surfaces in these cells. A potassium film which has been acted on by hydrogen, thus forming potassium hydride, is the one that seems to be the most used at present,

though there are other good ones. If the cell is to be employed in the measurement of ultra-violet light, the enclosing walls must be of fused quartz. If it is to measure light as we see it, a special filter is required to make its curve of sensitiveness match that of the eye; with this the cell can then measure amounts of light quantitatively, and thus act as a superior photometer. Very feeble photoelectric currents have been made measurable by using the methods of amplification so familiar in radio (p. 456).

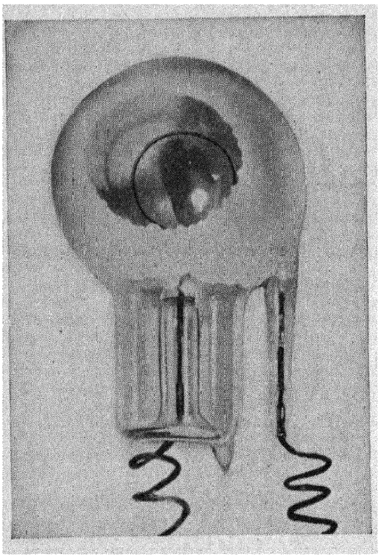


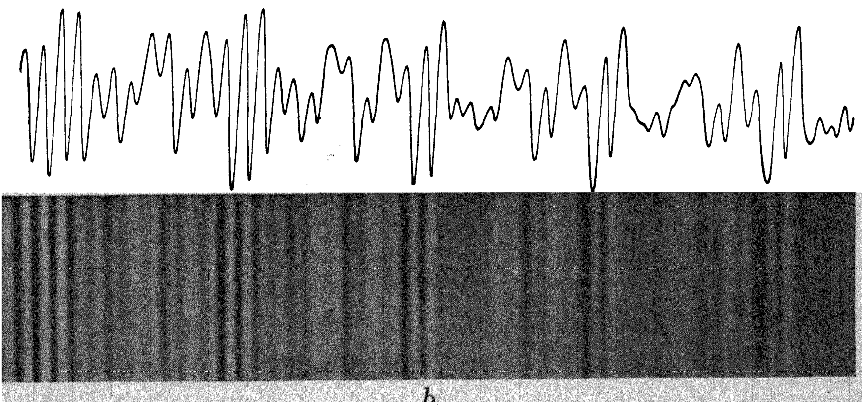
Fig. 38-2
A photoelectric cell. (Courtesy of the Bell Telephone Laboratories)

The electric circuit must include some instrument for measuring the current. A sensitive galvanometer is used if the current is not too small; otherwise, the rate of charging of an electrometer, or measuring electroscope, yields a value for the current. If the photoelectric current is fairly large, it may be used to control mechanical devices, instead of deflecting a galvanometer system; for instance, lights may be shut off at sunrise and turned on again at night by this means. It has been applied to the control of traffic lights, the detection and measurement of smoke in the air, the counting of passing automobiles, the running of a clock, the measurement of the area of hides, and to many other uses.

Practical applications. The ***talking motion picture*** and ***television*** are two recent practical developments which depend on the photoelectric cell. In the first of these a photographic record of the sound waves is made upon the film. This is accomplished by making the sound vibrations move one jaw of a narrow horizontal slit through which light is passing to the film. A loud sound opens the slit wide during each vibration and allows a large amount of light to reach the film. The film travels vertically while the record is being made. Figure 38–3*a* shows one of the pictures in such a strip, with an enlargement (*b*), of the sound "track," together with an accompanying curve exhibiting the sound waves in the usual form (traced by a "microphotometer").

To reproduce this record, light is sent upon the sound strip on the moving film through a horizontal slit and then falls on a photoelectric cell. The variations in the blackness of the deposit on the film produce similar variations in the light and therefore in the current in the cell. These variations are amplified in the manner already familiar (audio amplification), and are then led to a loud speaker. The motion of the film for picture projection must be jerky, but for sound vibrations continuous. As a result the photoelectric cell is placed below the projection apparatus and a foot or two of slack film lies

a



b

FIG. 38-3

between them. Thus the sound record beside a given picture is synchronous not with that picture but with one a little distance ahead or behind.

In the apparatus for ***television*** a small beam of light is made to travel back and forth over the object whose image is to be transmitted. It travels very fast in a crisscross fashion so as to cover the whole of that object a large number of times during each second. This process is called "scanning" the picture. In each position of the light beam it strikes upon a part of the "picture" whose reflecting power depends on how light or dark it is. A photoelectric cell near by catches some of this reflected light, and a

current is created in the cell circuit which varies exactly with the brightness of the light reflected from each spot. This current can be amplified, transmitted to a distant station and there made to throw a spot of light on a certain part of a screen which the observer is watching. The essential feature of the apparatus is to make the motion of the scanning beam synchronize with and be followed faithfully by the motion of the reproducing beam. Synchronous motors are run at both stations and maintained alike by radio, or by vibrations transmitted by wire; these motors control the motions of the two beams. The entire picture is completely scanned in 1/20 second, so that the observer sees it as a unit, his eye being incapable of following the motions of the spot of light because of its rapidity. The spot of light must, of course, be small enough so that the image is not obviously coarse-grained.

The transmission of pictures by telegraph is an essentially similar process except that a photographic film is substituted for the receiving screen, and the operation can be carried through in a more leisurely fashion. The picture to be transmitted is on a film wrapped into a cylindrical form, and this is rotated in front of the fixed "scanning" light, and fed forward by a screw at the same time.

Photoelectric theory of photographic action. When one takes a picture, a sensitive film is somehow altered, so that later on, when it is immersed in a certain solution, an image will become visible upon it. The change is invisible, and it is interesting to inquire into its nature. The subject is a large one, however, and we shall go no further with it than to say that the action seems to be a photoelectric one.[1] If an electron is displaced from a photographic "grain" in a film through the absorption of a quantum of light, that electron may merely take up a new position, the extra energy going into the form of potential energy associated with this displacement. There may be forces capable of holding the electron there indefinitely. If so, that grain is in one sense "ionized"; at least its constituent electric charges are separated more widely than usual, and it may therefore have a strong external electric field. This makes the initiation of chemical action much easier, so that this grain tends to "develop" (that is, metallic silver is formed in it and it turns black) while others not so affected remain unchanged. Thus the picture is formed.

This subject is one of great complexity. For instance, it has recently been found that the sensitiveness of a photographic film depends chiefly on the presence of an extremely small amount of an organic substance (allylthiourea) which in the past always managed to get into photographic films in the guise

[1] The reader interested in photoelectricity will find an interesting account of the whole subject in "Photoelectricity," by H. Stanley Allen, 2d ed., 1925 (Longmans, Green).

of an impurity in the gelatine. Just how this is able to encourage the photoelectric action appears to be still an open question.

Luminescence. A material which glows with a light of its own, not produced by a high temperature, is said to be ***luminescent.*** Many such cases are due to some sort of chemical action, for instance, the glowing of phosphorus. The light of the firefly, glow-worm, etc., can probably be explained on this basis also. Omitting such actions we shall consider two other somewhat similar phenomena which physicists call fluorescence and phosphorescence. A ***fluorescent*** material is one which glows with a light of its own when light, usually of a quite different color, falls upon it. A ***phosphorescent*** material (excluding cases due to chemical actions) is one that does the same thing but continues to glow after the exciting light has been shut off. There is no important difference between these two. Fluorescent materials also glow for a while after the stimulation is removed, but the time may be very short, perhaps of the order of a thousandth of a second only.

One can test materials for this effect by sending a concentrated beam of light (preferably blue) through a rapidly revolving disc, perforated with many holes, with the specimen mounted just behind them, and then viewing the specimen through a hole which is a little to one side of the one through which the light goes. By varying the angle at which one looks, one can choose to see the specimen while it is being illuminated or in the intervals between. By speeding up the disc the time intervals may be made extremely short. One's teeth and finger nails, and many common materials such as paper, are seen to be fluorescent to a slight extent, and substances like kerosene, machine oil, eosine, fluorescine, uranium glass, zinc sulphide, and many others show the effect strongly. The proper color to excite the fluorescence most strongly varies with the substance, but *the fluorescent light is of longer wave-length than the exciting light.* Thus if the substance glows with a green light, the chances are that the best exciting light is blue, while a blue fluorescence may require ultra-violet to excite it. A specimen of uranium glass lights up with a brilliant greenish-yellow glow when light from an arc lamp, filtered through cobalt-blue glass, is focused upon it, and the contrast is very striking, as there is no green or yellow in the light transmitted by the glass.

The fluorescence of many substances under the action of X-rays has already been mentioned. In these cases it is evident that the exciting waves are very much shorter in length than the fluorescent ones. The same substances are usually excited equally well by cathode rays.

Explanation of fluorescence and phosphorescence. The act of fluorescence occurs because short-wave light is absorbed by a

substance, presumably on account of the fact that the frequency of the incident light agrees with a natural frequency of vibration of the substance, or, in terms of the quantum theory, because the incident quanta are of the right size to excite the molecules of the substance and give them energy enough to reach some higher stationary state. The energy of this quantum might be disposed of in several ways. It might be turned into heat, used up in creating permanent changes in the material, re-emitted as it entered, or be used partly in one way and partly in another. Apparently in fluorescent materials some of the energy is lost, possibly transformed into heat. The rest goes into making a change in the nature of arrangement of the material, which persists for a very short time if the substance is fluorescent, and longer if it is phosphorescent. When its proper time comes, sooner or later the changed molecules revert to their original condition and re-emit the energy they had in storage. Since this was not quite the whole of the incident quantum, the light emitted by the substance is composed of smaller quanta, and is therefore of longer wave-length.

CHAPTER 39

RADIOACTIVITY [1]

Discovery of Becquerel rays, 605; the three sorts of rays, 606; isolation of radioactive materials, 607; atomic disintegration, 608; the nucleus of the atom, 609; radioactive transformation, 610; isotopes, 611; the age of the earth's crust, 611; counting individual atoms, 612; tracks of alpha and beta particles, 613; artificial disintegration of atoms, 615; structure of the nucleus, 616; penetrating or "cosmic" rays, 617; relativity, 618.

Discovery of Becquerel rays. At the time of the discovery of X-rays, there was some reason to suspect that these rays arose from the fluorescence of the glass walls of the tube which (in the earliest form of X-ray apparatus) emitted them. Accordingly Becquerel [2] in 1896 made a series of experiments on fluorescent and phosphorescent materials in the course of which he found that a quantity of a uranium salt (double sulphide of uranium and potassium) when laid over a photographic plate, carefully protected from light by a wrapping of black paper, and left there for a few hours, could produce an effect which was visible on the plate after it was developed (Fig. 39–1), just as light effects are. He later found that this would occur whether the substance was luminescent or not; that is, the new rays were spontaneously emitted from the uranium compound. Soon they were found to consist of three sorts of rays, which were named alpha (α), beta (β) and gamma (γ) rays. The α rays ionize the air near by and can easily be

FIG. 39–1

A photographic plate was wrapped in black paper, a key rested on the paper, and a piece of a uranium mineral ("pitchblende") was laid on top of the key and left for three days. This picture was the result.

[1] This chapter may be omitted in a short course.

[2] Antoine Henri Becquerel (1852–1908); French engineer and physicist; awarded the Nobel prize, jointly with P. Curie and his wife, for these discoveries; a member of a family distinguished for three generations for research in chemistry and physics.

detected through this action; they are stopped by thin sheets of aluminum, or even of paper. The β rays are more penetrating and affect a photographic plate; they ionize the air very little. The γ rays are the most penetrating of all, passing through a centimeter of lead or a hundred yards of air without serious diminution. Substances emitting these rays are said to be ***radioactive.***

The three sorts of rays. Experiments on these new rays were immediately undertaken in many laboratories, and it was soon found that a very limited number of substances produced them, but that they did so with complete indifference to their physical condition. Raising the temperature, for instance, or cooling with liquid air failed to alter the strength of the rays emitted by these substances. The source of all this activity was a mystery, and it looked as though a form of perpetual motion had been discovered. The clue to the puzzle was not found till the nature of the three sorts of rays had been determined.

The ***alpha rays*** were at first detected by their ionizing action on the air. A certain strength of them would be capable of creating a certain current between two charged plates in air, connected to a small battery. It was found, by allowing them to strike an insulated body connected to an electroscope, that they bore a positive charge. Later it was shown that they could be deflected, though with difficulty, by high electrostatic and magnetic fields, from which fact, as with cathode rays, it was possible to find their velocity and their ratio of charge to mass (e/m). Thus it was found that they were atoms four times as heavy as hydrogen atoms and moving at a speed (up to 2×10^9 cm./sec.) much greater than had ever before been known for anything so heavy as an atom. Later enough of these atoms were collected in a vacuum tube to make it possible to discover their nature through their spectrum, and they proved to be atoms of helium bearing (while acting as alpha rays) a double rather than a single charge. The normal helium atom has two electrons; the alpha particle has lost them both and is a naked helium nucleus, at least through the first part of its flight. Moving at such a tremendous speed, it breaks up most of the atoms directly in its path, and leaves a trail of ions behind it. Being so (relatively) large, however, it cannot pass through much solid matter, and hence the alpha rays are not very penetrating. They are stopped easily by a thin sheet of a heavy substance, though they will pass through glass in thick-

nesses less than 0.01 millimeter. Their range in air is a few centimeters; it is so sharply defined that alpha rays from different sources can be distinguished by means of it.

The ***beta rays,*** on the other hand, are roughly a hundred times more penetrating than the alpha rays and are easily deflected by magnetic fields. They carry a negative charge. They produce a feeble ionization in a gas through which they pass, breaking up about one per cent of the atoms they hit. The value of e/m for beta rays shows that they are electrons, but they move at prodigious speeds, up to within less than one per cent of the velocity of light itself. As there are theoretical reasons for believing that an electron would need an infinite amount of energy to make it move with the velocity of light, we see that the beta rays must arise from some sort of explosion of extreme violence.

The ***gamma rays*** prove to be identical with X-rays, but to have very short wave-lengths, down to 0.005 A.U. or one-millionth of the wave-length of green light. These rays are extremely penetrating; they ionize the air slightly and also produce fluorescence, and photographic effects, but they are usually faint and are completely unaffected by electric or magnetic fields.

Isolation of radioactive materials. In order to discover the sources of these rays, the active minerals were treated chemically and separated into various parts. In this way it was found that the activity went with certain substances. In the course of such work P. Curie and his wife, and later Mme Curie alone,[1] and others isolated a number of substances, the best-known of which is radium, the element with atomic number 88, and chemical properties and spectrum like barium. Radium, though not the most active element, is more than a million times more active than uranium, with which it is always associated. It occurs in such minute amounts that many tons of uranium ore are required to produce one gram of radium chloride. Thus it has the distinction of being one of the most expensive materials on the market.

[1] Pierre Curie (1859–1906), professor in Paris, already distinguished for his experimental work on piezo-electricity and magnetism before he began research on radioactivity. Marie Sklodowska Curie succeeded him in his professorship, and greatly extended their discoveries. She is now probably the most distinguished woman engaged in scientific research. Since her pioneer work described here she has devoted a great deal of attention to the problem of relieving cancer, etc. by radioactive treatment. In 1929 she was given the sum of $50,000 by President Hoover as a second American gift for the purchase of radium.

Many other very active materials were isolated. Mme Curie named the first one that was discovered polonium, after her native country; this one arises in small amounts from the disintegration of radium itself. A table of these materials is given below (p. 610). Several of these are active enough to be able to maintain themselves at a temperature slightly above their surroundings, mainly through absorption of their own alpha rays. The search for the source of this energy led Rutherford[1] to propose the theory of atomic disintegration, which at the same time accounts for the scarcity of radioactive materials and for their properties.

Atomic disintegration. As one goes down the list of the known atoms, one finds heavier and more complex ones and they might reasonably be supposed to reach a limit beyond which they would become unstable and could no longer exist. Even the heaviest ones that do exist might well be so nearly unstable that some slight accident might upset them, and then they might explode, ejecting portions of their own structure and settling down as simpler, more stable, and somewhat lighter atoms. This idea seems to be in accord with the facts; only the heaviest atoms appear to be appreciably radioactive, and through the study of these a whole series of new elements have been found, the atoms of which have only a temporary existence. It seems to be pure chance, or at least something unknown governed by the laws of pure chance, that dictates when they will explode. If a given quantity of a certain radioactive material is on hand, half of it will have been transformed into something else in a certain time, known as the "half-life." Half of the remainder will go in the next interval of the same length, and so on. This law of decay is an exponential one. If M is the mass of the material that is transformed in a time t, M_0 the original mass (at the time $t = 0$), the equation of decay is

$$M = M_0 \varepsilon^{-\lambda t}$$

where ε is the Napierian base of logarithms (2.7183) and λ is the "decay constant." The "half-life" is then given by finding the value of t in this equation when $M = \frac{1}{2}M_0$; that is,

[1] Sir Ernest Rutherford, professor of physics at the Cavendish Laboratory, University of Cambridge, England. At the time referred to he was professor in McGill University, Montreal. He received the Nobel prize for physics in 1908, and has written admirable accounts of this field of research in "Radioactive Transformations," 1906 (Scribner's) and "Radioactivity," second edition, 1905 (Cambridge University Press.)

$$t = \frac{1}{\lambda} \times \log_e 2 = \frac{0.69}{\lambda}.$$

Table XXXI (p. 610) gives the values of the half-life for several active elements.

The nucleus of the atom. Various reasons point to the nucleus as the seat of the instability of the atom. We know from X-ray spectra that there is nothing in the atomic structure external to the nucleus which could be as heavy as an alpha particle. Moreover, the most energetic change in any electron structure, that responsible for the emission of the so-called *K* X-ray line in the spectrum of uranium, corresponds to a wave-length of about 0.1. A.U., whereas there are gamma rays one-twentieth as long, or containing twenty times as much energy. Such a supply of energy cannot be supposed to come from the external electrons.

Then, also, Rutherford in 1911 obtained an estimate of the diameter of the nucleus which shows that matter is packed more closely there than anywhere else. If the electrons could be knocked off from all the atoms in a material, and the remaining nuclei placed close together, we should obtain matter some 10,000 times denser than usual. Eddington[1] has shown that this condition probably exists in certain stars which seem to have unusual properties. Rutherford's estimates of the size of the nucleus were obtained from experiments on the scattering of alpha particles by thin films of gold, etc. These showed that when the alpha particle collided with an atom, there was an unexpectedly high probability that it would be deflected through a large angle. Such a deflection requires the existence of very great forces that are called into being during the collisions, and these can occur only if the nucleus is concentrated into a very small space, permitting the alpha particle to come very close to it and yet still remain outside it. The estimated diameter of the nucleus of the gold atom was of the order of 10^{-12} centimeters, or ten thousand times less than the width of the atom itself; and in this small space must be concentrated practically all the mass of the atom. We shall see that we now have ways of getting some information about the state of affairs within even this minute sphere.

[1] A. S. Eddington, professor of astronomy at the University of Cambridge; a scientific writer with a delightful style; author of "Stars and Atoms," 1927 (Yale University Press), which everybody should read who has any taste for science; also "The Nature of the Physical World," 1928 (Macmillan), "Space, Time and Gravitation," 1920 (Cambridge University Press), and others.

Radioactive transformation. Beginning with the heaviest element, uranium, Rutherford and many others have discovered the manner of its disintegration, the succession of elements that are thus formed, their "half-life," their atomic weight and atomic numbers, and something of their chemical properties. In Table XXXI the two main series of radioactive elements (those starting with uranium and thorium) are given together with some of their properties. There is in addition a third series, not given here, probably derived as an alternative; uranium II may have two possible changes, the one to ionium shown in the table of the uranium-radium series, and another occurring less frequently, to uranium Y and hence to the actinium series (omitted from the table).

TABLE XXXI

Name	Half-life	Rays emitted on disintegrating	Atomic Weight	Atomic Number	Range α rays (in cms.)	Velocity α rays (10^9 cm./sec.)
Uranium	4.5×10^9 years	α	238	92	2.83	1.42
Uranium X_1........	23.8 days	β	234	90	....	....
Uranium X_2........	1.15 minutes	β	234	91	....	....
Uranium II........	2.10^6 years	α	234	92	2.91	1.43
Ionium............	10^5 years	α	230	90	....	....
Radium...........	1580 years	α	226	88	3.39	1.51
Radium Emanation or Niton.........	3.81 days	α	222	86	4.12	1.61
Radium A.........	3.05 minutes	α	218	84	4.72	1.69
Radium B.........	26.8 minutes	β	214	82	....	....
Radium C.........	19.5 minutes	β	214	83	....	....
Radium C′.........	10^{-7} seconds	α	214	84	6.97	1.92
Radium D.........	16 years	β	210	82	....	....
Radium E.........	4.85 days	β	210	83	....	....
Radium F (Polonium)	136 days	α	210	84	3.92	1.59
Lead..............			206	82	....	....
Thorium...........	2×10^{10} years	α	232	90	....	....
Mesothorium I.....	6.7 years	β	228	88	....	....
Mesothorium II....	6.2 hours	β	228	89	....	....
Radiothorium......	1.90 years	α	228	90	....	....
Thorium X........	3.64 days	α	224	88	....	....
Thorium emanation.	55 seconds	α	220	86	....	....
Thorium A.........	0.14 seconds	α	216	84	....	....
Thorium B.........	10.6 hours	β	212	82	....	....
Thorium C.........	61 minutes	β	212	83	....	....
Thorium C′........	10^{-11} seconds	α	212	84	....	....
Lead..............			208	82	....	....

Some interesting things may be noted about this table. One is that the final product is in both series the familiar metal, lead, which

is not radioactive, and has an ordinary atomic weight of 207.2. But lead derived from thorium should have an atomic weight of 208, and this conclusion is verified when lead is tested which has been obtained from thorium minerals free from uranium. On the other hand lead derived from uranium is found to have an atomic weight of nearly 206, as it should according to this theory of its origin. A mixture of two sorts of lead in proper proportions yields the usual atomic weight.

Isotopes. These two sorts of lead are isotopes (p. 448). They each contain eighty-two electrons, arranged in identical fashion, so that their chemical properties are the same, but their atomic weights are different. Other isotopes are to be found among the radioactive elements in Table XXXI. For instance, uranium X_1 is an isotope of thorium; radium of thorium X, radium B of radium D, etc. Unfortunately many of the radioactive isotopes have a rather transient existence and therefore can be obtained only in such extremely small amounts that experiments of a chemical sort can hardly be performed upon them. But among ordinary substances that occur in the form of isotopes there are others, besides lead, that can be obtained in large quantities and partially separated out into their isotopic constituents. This can be done by any process in which the difference in mass might alter the behavior of the atoms. Diffusion and evaporation, for instance, are two actions which might go on slightly faster with the lighter atoms than the heavy. By successive evaporations two small specimens of mercury have been prepared which differed from each other in density by a few hundredths of one per cent. Mercury is known from positive ray experiments to consist of a mixture of seven isotopes with atomic weights of 196, 198, 199, 200, 201, 202, and 204. The separation, though not by any means complete or satisfactory, at least gives independent evidence in support of the existence of isotopes.

The mass spectrograph (p. 448) gives much more definite evidence.

The age of the earth's crust. As one of the many possible deductions from the facts of radioactivity, it is interesting to note some conclusions that have been reached in regard to the age of the minerals containing uranium; that is, presumably, the time since they assumed their present solid form. If all the lead in a uranium mineral comes from the decay of the uranium and the various intermediate products, and if the half-life of all these materials is known, the rate of growth of the lead can readily be calculated. The proportion of lead

to uranium in the material will evidently increase with time. If the proportion is measured chemically, the time becomes known. These estimates lead to ages of over 1,000,000,000 years, for at least this portion of the earth's crust.

Counting individual atoms. Duane, Geiger and others have developed an ingenious scheme whereby one can detect the arrival in a testing vessel of a ***single*** alpha particle, or even a single X-ray quantum. The apparatus (Fig. 39–2) consists of a small, metal, cylindrical box in a partial vacuum. Down the middle of the box passes a fine wire, insulated from it. Such a difference of potential is maintained between the wire and the cylinder that a discharge will just ***not*** pass between them. When an alpha particle is shot from the source S through the small hole, it creates a large number of ions near the wire and these cause a momentary current to flow, large enough to deflect a galvanometer. Thus Rutherford and Geiger were able to count the number of alpha particles emitted per second from any radioactive preparation and entering the hole. From this, and the area of the hole, they could find the total number of particles emitted per second in all directions. They were thus enabled to calculate the amount of helium created from a given amount of radium in a given time. It was small, but large enough to indicate the possibility of detecting it spectroscopically, and measuring its volume, which, as already mentioned, Rutherford actually succeeded in doing. As a result, the number of gas particles in a cubic centimeter was obtained and the value obtained from the kinetic theory was thus independently verified.

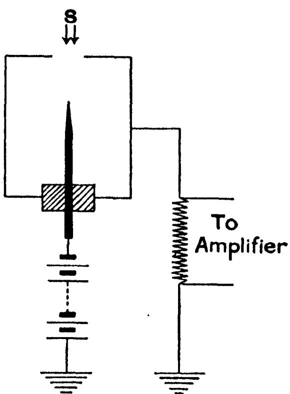


Fig. 39–2
A point counter

In recent years the counting method has been improved. The newest "point counter" uses a sharp steel point instead of a wire and the sudden rush of current from the arrival of each particle is amplified by radio methods so that each alpha particle makes a thump in a loud speaker audible to a large audience; or it can be made to record its arrival automatically by the motion of a pen on a piece of paper.

If a small area covered with crystals of zinc sulphide is exposed to a bombardment by alpha particles, it is seen to glow in the dark.

This is the device commonly used to make watch dials luminous. If one takes a ten-power magnifier to examine one of these luminous surfaces (having first rested one's eyes in the dark for ten minutes), it will be seen that this luminosity consists of a multitude of minute splashes of light occurring at random all over the surface. In this case the radioactive material is mixed with the crystalline material and the latter furnishes the light when struck by the alpha particles ejected by the atomic explosions. By means of a "counter" of the type first described, Rutherford showed that ***each alpha particle*** hitting such a crystalline material ***produces a splash of light.*** Hence by watching such a surface, the number of alpha particles reaching it can be counted. Naturally this method, though simple, is fatiguing and is not now used if a good "counter" is available.

Tracks of alpha and beta particles. Wilson [1] in 1912 devised a beautiful method whereby the tracks of individual alpha or beta particles can be photographed and studied at leisure. It had long been known that when air saturated with water vapor is suddenly expanded, the resulting cooling causes a cloud of small drops to form which may persist for some time. Such drops form more easily on something than they do by themselves, and they seize upon dust particles if there are any, to serve as centers upon which to begin condensation. The first drops that form on dust particles, if allowed to settle, will carry these down and leave a dust-free air for future experiments. In such air it is very difficult for a cloud to form, but if there are ions present, they act as centers of condensation as readily as dust particles. Wilson arranged a chamber in which the air was kept moist, and a sudden expansion could be produced in it. When alpha rays from a speck of radioactive material were shot into this air just as this condensation occurred, each particle created a line of ions along its track on which small water drops were formed. This track was visible as a white line for a moment, before the individual drops had time to diffuse away and evaporate, and it could then be photographed. Such tracks are shown in Fig. 39–3. The alpha-ray tracks come to a sudden end when the velocity of the particles reaches a value too low to cause ionization. Their uniformity in length shows that the alpha rays have a definite range (in air), and this indicates that the

[1] C. T. R. Wilson, of Cambridge, England, who in 1927 shared the Nobel prize for physics with A. H. Compton.

particles are all ejected at the same speed. Near the end of a track a sudden turn is sometimes seen. In such a case the particle must have made a direct hit on an atomic nucleus. In rare cases (as in Fig. 39–3) a track may be forked. When this happens it appears likely that the alpha particle is captured by the nucleus, while fragments of the latter are thrown out, as by an explosion, forward or backward. Such cases are considered below.

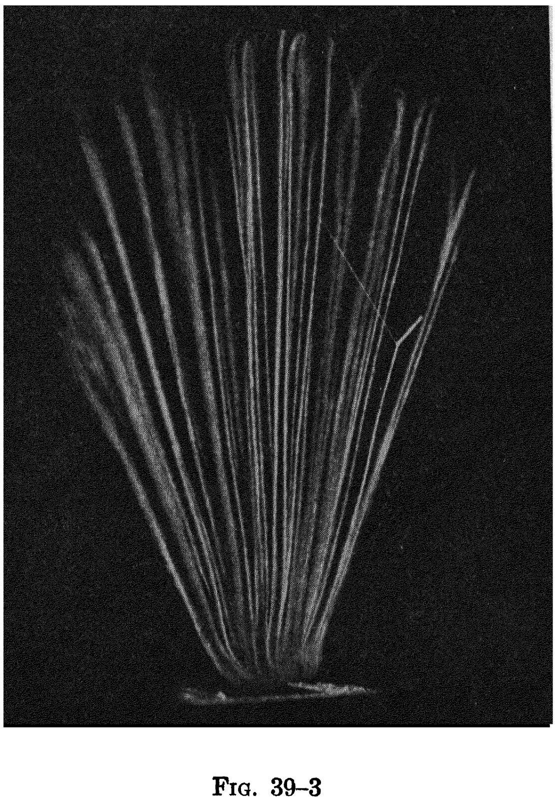


Fig. 39–3
Alpha-ray tracks in nitrogen, photographed by Professor W. D. Harkins, University of Chicago

Beta rays also make tracks, but these are not so easily followed, as the moving electron evidently passes through a very great many atoms without disturbing them. Its high speed seems to carry it through an atom before the disruptive forces which it produces have time to act. Its path in Wilson's apparatus is marked by occasional drops, and seems often to be curiously curved, as has been shown by stereoscopic photographs.

X-rays when shot into Wilson's apparatus ionize the air, and the electrons ejected from the atoms create little crooked tracks (Fig. 39–4) which are better marked than those of the β rays, since the velocities of the electrons are less and this enables them to ionize the atoms they strike more frequently. These can be called photoelectrons, as the process of ejection is similar to that by which ordinary light acts, except that the electrons come from deeper layers in the electronic arrangement of the atom.

Artificial disintegration of atoms. Rutherford has tried experiments in which alpha particles have been sent through a mass of gas, and the gas has been examined to see whether any fragments of atoms seemed to be flying off as a result of impacts of the kind mentioned above. The range of the alpha particles from a radio-

FIG. 39–4

Trail of an electron ejected photoelectrically by X-rays in air (after Wilson)

active material is very definite; for instance, those from radium C go about seven centimeters in air and then stop. But Rutherford found that in nitrogen an occasional particle went very much farther, whose behavior could be explained only by supposing it to be a light fragment of the nitrogen nucleus shot out as a result of a direct hit. Figure 39–3 shows in the case of the forked track that one of the lines is unusually thin and long. This could be explained as due to a particle which was lighter than an alpha particle, and thus produced fewer ions as it rushed along. Rutherford was able to show that such particles are hydrogen nuclei ("protons") shot out from the nucleus of nitrogen. Thus an atom of hydrogen is created from one of nitrogen. The nucleus of the latter may capture the bombarding alpha particle, and thus become a nucleus of atomic weight $14 + 4 - 1$ or 17, an isotope of oxygen. This kind of transmutation of the elements

has been found only with elements of odd atomic number, such as boron, fluorine, sodium, aluminum, etc. No such effects were obtained from oxygen, or other even-numbered elements.

Structure of the nucleus. The extraordinary stability of the helium nucleus, as shown by its career as an alpha particle, leads one to suppose that the nuclei of other very stable atoms are made up of alpha particles, at least as far as possible, with a few protons added to make up the necessary weight. Oxygen, for instance, has a weight equal to that of four alpha particles; carbon of three; nitrogen needs three alpha particles and two protons. It is observed that those elements whose atomic weights are multiples of that of helium include most of the commonest materials in existence, such as carbon, oxygen, magnesium, silicon, sulphur, calcium, etc. If we add the inert gases, neon and argon, which satisfy the same condition, we may conclude that a nuclear structure composed of alpha particles is particularly well built and stable. This is in harmony with Rutherford's results mentioned in the last paragraph.

The work on isotopes (p. 448) shows that all atoms have whole numbers for their atomic weights; those that appear not to satisfy this condition are really mixtures of whole-number isotopes. This strongly supports the conclusion that the nuclei of all atoms are made up of alpha particles and protons. How these can be held together is as yet somewhat mysterious. Their charges would produce strong electrostatic repulsions, unless the electrostatic laws break down at such small distances. The existence of high-speed beta rays from radioactive materials points to the conclusion that there are electrons also packed into their nuclei, and these would be useful in helping overcome the electrostatic repulsions of the alpha particles. One has to be a little cautious in reasoning that since electrons are shot out of the nucleus, they must exist as such within. It was a wise scientist who once remarked that because smoke comes out of a cigarette, it does not follow that the cigarette is made of smoke.

We might well consider how the nucleus of an atom is composed. If we take ***silicon*** as a first example, we see that it has an atomic number of 14 and an atomic weight of 28. There should be seven alpha particles crammed into its nucleus, in order to make up its weight, each of which presumably has its positive charge of $+ 2e$, or $+ 14e$ in all. The outer structure of fourteen electrons specified

by its atomic number just balances this and leaves the atom neutral. Considering ***tungsten*** as a somewhat more complicated case, we find that it has an atomic weight of 184, and its nucleus could thus be made up of forty-six alpha particles. These would then bear a charge of $+ 92e$, but the tungsten atom has an atomic number of only 74, and therefore a net nuclear charge of $+ 74e$. This requires the obliteration of a charge of $+ 18e$ in the nucleus, which we must imagine to occur from packing eighteen electrons into it. Similarly, the nucleus of the ***arsenic*** atom can be seen to contain eighteen alpha particles, three protons, and six electrons to make up its atomic weight of 75 and atomic number of 33. The alpha particle itself, if made up of protons, must contain four of these and also two electrons, in order to keep its charge down to $+ 2e$.

Evidence as to the internal structure of the nucleus is obtainable in several ways. Rutherford has recently made some suggestions of a definite sort as to how the alpha particles, protons, and electrons may be arranged inside the nucleus in the case of certain elements, but this part of the subject is still in an incomplete stage, and we shall not pursue it further.

Penetrating or " cosmic " rays. In 1903 Rutherford discovered that a few ions can always be found in the air on account of the existence of a very penetrating type of radiation, which is much less absorbed by matter than X-rays or even than gamma rays, which were up to that time the most penetrating ones known. Many experimenters have since made measurements on these rays, sifting them out from other less penetrating types by shielding the measuring electroscope with lead walls, or testing them down in mines, up on the tops of mountains, over surfaces of ice, and so forth. Millikan in recent years has made especially interesting experiments with instruments carried to great heights (10 miles) by balloons, or sunk in the water of high mountain lakes. As a result of all this work he concluded that these rays can penetrate through six feet of lead or sixty-eight feet of water before they become too faint to detect; that they originate outside the earth and are more prominent the higher up we go; and that they do not come from the sun. He gave them the name "cosmic" rays, because his experiments indicated that they arose in the heavens, either in distant stars, or through some process occurring in "empty" interstellar space. Astronomers now have reason to believe that the whole of this empty space contains at least one molecule per cubic centimeter, which seems very little until one tries to think of the number of cubic centimeters involved, when the total amount of matter in space is seen to be huge. Millikan imagined a process of atom-building to be going on throughout these vast regions. As a part of this radical assumption, he supposes a large quantum of left-over energy to be sent out each time an atom is made and uses a theoretical method to calculate the energy involved and the wave-length of the

resulting radiation. As the theory is perhaps not entirely settled, and as there is not yet complete agreement as to the nature of these radiations, the results may not remain in their present form, but in any case they are worthy of consideration on account of the boldness and originality of the conceptions through which they were obtained.

Relativity. The theory of relativity, due largely to Einstein, deals with the effects on the laws and concepts of physics produced by the motion of bodies. While the changes thus brought about are almost immeasurably small except for speeds approaching that of light, the theory has led to a reorganization of our ideas, and we must consider it briefly.

We look on our units of length, mass, and time as unchanging things suitable for the most exacting measurements; and so they are ordinarily. But, let one observer establish a laboratory similar to ours on a moving train, and send us radio time signals. It is easy to see that his signals arrive late, as do ours to him also, on account of the time taken by the radio waves in traveling. But, even after this has been allowed for, it can be shown that our clock seems to him to be running slow, and we say the same about his. Thus his unit of *time* and ours no longer agree.

By somewhat similar reasoning it can be proved that his unit of *length* also has changed, and is no longer the same as ours. Thus we might say that his motion spoils his measurements, while we remain correct because we are at rest. But astronomers remind us that the earth rotates about its axis, revolves about the sun, and is rushing with the whole solar system through space; thus we are moving too. If motion alters our units, who is right? Evidently no one can claim to be, and we must find, if we can, some system which will yield invariable laws, unaltered by motion, if we are to have scientific results of universal validity.

Einstein in 1905 proposed his "special theory," dealing only with uniform velocities. He suggested that if something happens at a given point, (having three space coördinates, x, y, and z, to define its position) at a certain time, t, we might refer to the *event* as having four coördinates (x, y, z, t), or as having a certain "position" in four-dimensional "space-time." We can compute the space-time "interval" between two events, and *this can be made the same for all observers*, (by choosing a proper definition of what we mean by an interval) and therefore absolute, independent of the motion of any observer. This may then be made the basis of a whole new system of geometry, from which follows a new system of mechanics also. A body free from force moves in a straight line, and its path in space-time is the analogue of a straight line, though, owing to the curious properties of this space-time, the straight line is no longer the *shortest* distance between two points (events) but the *longest*. We may call such a line the *natural path* of such a body.

In his "general theory of relativity" (1915) Einstein imagined that a large mass produces a curving or warping of space-time near it, so that a particle traveling on its natural path now follows a curve in the warped region, somewhat as a walker following the easiest path across a plain will curve his track when he comes to a hill, neither going straight up and over it, nor altogether avoiding it by a long detour. This furnished an "explanation" of gravitation in terms of the properties of space and time near a large mass, and led to a

prediction that a ray of light would be deviated by coming near to a heavy body. This has been verified at various solar eclipses by finding a small apparent displacement of the positions of stars seen near the sun.

The theory also indicates that a body which acquires kinetic energy thereby increases its inertia in the direction of motion. The idea is familiar (p. 445) in the case of an electron from the usual electromagnetic theory, but here it is generalized to apply to all bodies. The planet Mercury, which is nearest to the sun, moves in an elliptical orbit with variations of speed. Newton's theory of gravitation has failed to explain a slight anomaly in its motion, but Einstein showed that the change in mass due to its change in speed would bring theory and observation into agreement. One consequence of this part of the theory is that mass and energy are interchangeable. A star with more energy is thereby more massive. Our sun can supply us with heat by diminishing in mass. This theory is now held by many astronomers to explain the very long life of hot stars.

Another prediction of the theory was that the atomic mechanism which sends out light vibrations would be altered near a very massive body, and it has in fact been found that atoms on the sun do send out slightly slower vibrations than those on the earth, though the effect is excessively small.

Thus the theory in its general form destroys our conception of the constancy of mass and of energy, and yet it has been verified in cases where the dynamics of Newton has failed to yield the correct result, and it has predicted unexpected phenomena which proved to be true. What the future of this theory may yield no one knows, but we must unite in admiration of the intellectual feat involved in its conception and development.

Books recommended:

J. A. Crowther, "Ions, Electrons and Ionizing Radiations," 4th edition, 1924 (Longmans, Green).

Hevesy and Paneth, "A Manual of Radioactivity," 1926 (Oxford University Press).

A. S. Eddington, "Space, Time and Gravitation," 1920 (Cambridge University Press).

APPENDIX

Trigonometry Needed for This Book

Let BAC, Fig. A–1 be a right-angled triangle including the angle θ.

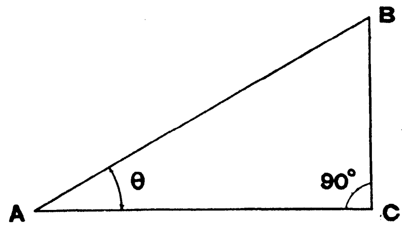


Fig. A–1

(*a*) The sine of θ (written $\sin\theta$) is BC/AB, i.e., the opposite side divided by the hypotenuse.

The cosine of θ (written $\cos\theta$) is AC/AB, the adjacent side divided by the hypotenuse.

The tangent of θ (written $\tan\theta$) is $BC/AC = \sin\theta/\cos\theta$.

(*b*) $\sin\theta = \cos ABC = \cos(90° - \theta)$

(*c*) $\overline{BC}^2 + \overline{AC}^2 = \overline{AB}^2$

Therefore $$(BC/AB)^2 + (AC/AB)^2 = 1$$

or $$\sin^2\theta + \cos^2\theta = 1$$

Whence

$$\sin^2\theta = 1 - \cos^2\theta \quad \text{or} \quad \cos^2\theta = 1 - \sin^2\theta$$

(*d*)

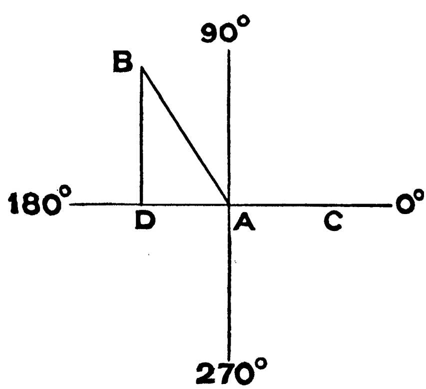


Fig. A–2

In case the angle θ is between 90° and 180° (Fig. A–2)

$$\sin\theta = \sin BAC = BD/AB = \sin BAD = \sin(180° - \theta)$$
$$\cos\theta = \cos BAC = -DA/AB = -\cos BAD = -\cos(180° - \theta)$$

The distance DA is taken negative because it is to the left of A. In case the angle θ is between 180° and 270°,

sin θ and cos θ are both negative.

$$\sin\theta = -\sin(\theta - 180°)$$
$$\cos\theta = -\cos(\theta - 180°)$$

In case the angle is between 270° and 360°,

$$\sin\theta = -\sin(360° - \theta)$$
$$\cos\theta = +\cos(360° - \theta)$$

Useful Numerical Data

Length

1 inch	= 2.5400 cm.	1 cm.	= 0.3937 inch
1 foot	= 30.48 "	1 meter	= 39.37 inch
1 yard	= 91.44 "		= 3.281 feet
1 mile	= 1.6093 km.	1 km.	= 0.6214 mile
1 mil	= 0.001 inch	1 micron	= 0.001 mm.

Volume

1 in.^{3}	= 16.387 cm.^{3}	1 cm.^{3}	= 0.0610 in.^{3}
1 ft.^{3}	= 28.317 liter	1 liter	= 1000 cm.^{3}
1 pint	= 568 cm.^{3}		= 1.76 pints
1 mi.^{3}	= 4.168×10^9 cm.^{3}		= 61.0 in.^{3}

Mass

1 pound	= 453.6 gram.	1 kg.	= 2.2046 pound

Miscellaneous

1 radian = 57°.296

π = 3.14159

π^2 = 9.8696

e = 2.71828

The Periodic Table of the Elements

0	1	2	3	4	5	6	7	8		
	1 H 1.0078									
2 4 He 4.002	**3** 7 6 Li 6.940	**4** 9 Be 9.02	**5** 11 10 B 10.82	**6** 12 13 C 12.00	**7** 14 N 14.008	**8** 16 18 17 O 16.00	**9** 19 F 19.00			
10 20 22 Ne 20.183	**11** 23 Na 22.997	**12** 24 25 26 28 Mg 24.32	**13** 27 Al 26.97	**14** 28 29 30 Si 28.06	**15** 31 P 31.02	**16** 32 34 33 S 32.06	**17** 35 37 Cl 35.457			
18 40 36 A 39.94	**19** 39 41 K 39.10	**20** 40 44 Ca 40.07	**21** 45 Sc 45.10	**22** 48 Ti 47.90	**23** 51 V 50.96	**24** 52 Cr 52.01	**25** 55 Mn 54.93	**26** 56 54 Fe 55.84	**27** 59 Co 58.94	**28** 58 60 Ni 58.69
	29 63 65 Cu 63.57	**30** 64 66 Zn 68 70 65.38	**31** 69 71 Ga 69.72	**32** 74 72 70 Ge 72.60	**33** 75 As 74.93	**34** 80 78 76 Se 82 etc. 79.2	**35** 79 81 Br 79.92			
36 84 86 82 83 etc. Kr 82.9	**37** 85 87 Rb 85.44	**38** 88 86 Sr 87.63	**39** 89 Yt 88.92	**40** 90 94 92 Zr 91.22	**41** Cb 93.1	**42** Mo 96.0	**43** Ma 98	**44** Ru 101.7	**45** Rh 102.91	**46** Pd 106.7
	47 107 109 Ag 107.88	**48** 110 111 Cd 112 113 etc. 112.41	**49** 115 In 114.8	**50** 120 118 Sn 116 124 etc. 118.7	**51** 121 123 Sb 121.77	**52** 128 130 Te 126 127.5	**53** 127 I 126.93			
54 129 132 131 134 etc. Xe 130.2	**55** 133 Cs 132.81	**56** 138 136 Ba 137.36	**57–71** Rare Earths 138 to 175	**72** Hf 178.6	**73** Ta 181.5	**74** W 184.0	**75** Re 188.7	**76** Os 190.8	**77** Ir 193.1	**78** Pt 195.23
	79 Au 197.2	**80** 202 200 Hg 199 198 etc. 200.61	**81** Tl 204.4	**82** 206 207 Pb 208 207.2	**83** 209 Bi 209.0	**84** Po 210	**85**			
86 Nt 222	**87**	**88** Ra 225.97	**89** Ac 229	**90** Th 232.12	**91** Pa 234	**92** U 238.14				

Rare Earths

57 La	65 Tb
58 Ce	66 Dy
59 Pr	67 Ho
60 Nd	68 Er
61 Il	69 Tu
62 Sa	70 Yb
63 Eu	71 Lu
64 Gd	

Note. In each space stands the chemical symbol of the element, its atomic number (upper left, in heavy type), its atomic weight (lower left), and its isotopes (on the right) when known, in the order of their abundance.

INDEX

ANSWERS TO PROBLEMS

Page 7

1. Spacing inversely as depth. **3.** 33.3 lbs. force. **4.** 39.2 lbs./in.^{2} **5.** 1500 grams; 1545 grams. **6.** 1 kg. force. **7.** 1 kg. **9.** 6.18 kg. **10.** 1108 lbs.; same. **11.** 86.6 lbs./in.^{2} **12.** 2 cm. below; same. **13.** 4.76 lbs./in.^{2}; 16,835 lbs.

Page 12

1. Gold, 23.5 cm.^{3}, 1.435 in.^{3} Iron, 57.5 and 3.51. Wood, 908–1300 and 55.4–79.3. **2.** (*a*) 1.333; 0.0214; (*b*) 40; 0.64; (*c*) 450; 7.2. **3.** 35.6 cm. square. **4.** 60 kg. **6.** Same density as water. **7.** 2.56 ft.^{3} **8.** 60 cm.; 66.67 cm. **9.** 19.89 lbs./in.^{2}; 25.3 lbs. **10.** 249.6 lbs. **11.** 1176 cub. m. **12.** 5 cm.^{3} **13.** 12,450 ft.^{2} **14.** 1411 lbs. **15.** 16,000 cm.^{3} **16.** 0.574. **17.** 1.1 grams/cm.^{3} **18.** 3.09 lbs. net force downward; tension at rest 6 lbs. **19.** 17.8. **20.** 0.046 gram, 0.13 gram, 0.08 gram. **21.** 400 grams; same. **22.** 11.45 grams. **23.** 119,786 kg.

Page 25

1. 54.4 cm. **2.** 4090 grams/cm.^{2} **3.** 47 lbs./in.^{2}; 152.4 in. height; 5.08 atm. **4.** 10.58 tons. **5.** In. **6.** No difference in any case. **7.** 0.00125 gram/cm.^{3} **9.** 44.7 lbs./in.^{2}; 3.04 atm.; 91.2 in.

Page 40

2. Vertical. **3.** F = 5.77 lbs. Tension 11.55 lbs. **4.** 28.9 lbs. **5.** 12.5 lbs.; 27.5 lbs. **6.** 43.3 lbs. **7.** 60 lbs.; 187.4 lbs. **8.** 150 lbs.; 250 lbs. **9.** 30 lbs.; 190 lbs. **10.** 125 lbs. **11.** 243.9 lbs.-ft.; 24.5 lbs. **12.** F = 57.7 lbs.; force at E = 49.3 lbs.; the man's upper arm cannot be quite vertical. **13.** 40.4 lbs. **14.** 41.6 lbs. **15.** 500 lbs. **16.** 76.4 lbs. **17.** 111.8 lbs.; about 26°.5 to horizontal. **18.** 30.8 lbs. on two; 1.67 on others. **20.** 59.5 tons; 53.5 tons.

Page 44

1. The first by 1 min. 17 sec. **2.** 2 mi./hr. **3.** (*a*) about 26°.5 down; 6.71 mi./hr. (*b*) 30° up; 5.2 mi./hr. **4.** 37°.18′ (with the bank). **5.** 17.3 ft./sec. **6.** 1.25 ft. **7.** 30 mi./hr. North. North-east. **8.** Each 42.4 mi./hr. **9.** 15 ft./sec. **10.** Angle of 22°.3 with line across track. **11.** 10.41 ft./sec.

Page 48

1. 163.8 ft. **2.** 64 ft.; 64 ft./sec. **3.** 1; 4; 9 ft.; 8; 16; 24 ft./sec. **4.** 4 ft.; 1 sec. **5.** 48 ft. up. **6.** 4590 m. **7.** 1132.5 m. **8.** 2010 m. **9.** 40 sec.; 80

ft./sec.; 1600 ft. **10.** 5 sec. **11.** 48 ft./sec.; 1.27 sec.; 48.6 ft./sec. **12.** 2.6 sec. **13.** $a = 6.4$ ft./sec.^{2}; 320 ft. **14.** $a = 2$ ft./sec.^{2}; 25 ft.; 7.07 sec. **15.** Rises 0.5 sec., or 4 ft.; falls 3 sec.; 144 ft.

Page 53

1. 100 ft. **2.** 327 ft./sec. **3.** 144 ft.; 30 ft.; 96.5 ft./sec. **4.** 0.5 sec.; same; vertical straight line; straight line inclined backward. **5.** 10 sec.; straight over it; 900 ft. **6.** 44.7 ft./sec. **7.** Falls 0.397 inch. **8.** 511.3 cm./sec. **9.** Straight line inclined backward; curved line forward. **10.** v^2/g. **11.** 3200 ft./sec.; 141.2 sec.; 80,000 ft., or 15.1 miles.

Page 56

1. 2.45 cm./sec.^{2} **2.** In a straight line, inclined outward; parabola to outside observer; pendulum at angle whose tangent is 2/49. **3.** 263 cm./sec. **4.** $v^2 = lg/\sqrt{2}$. **5.** 5.05 ft.

Page 68

1. Yes. **2.** The flea. **3.** *b* and *c* wrong. **5.** *b* wrong. **7.** 1 ft./sec.^{2}; 200 ft.; 6.25 lbs. **8.** 14.06 lbs. **9.** 108.9 cm./sec.^{2}; 222 gr. **10.** Go up together. **11.** 18.75 tons; 3.13 tons. **12.** 4.8 ft./sec.; same (but the work done is quite different). **13.** 125 ft. **14.** 250 lbs.; 1250 lbs. **15.** 510 grams. **16.** 400 lbs. **17.** 9000 lbs. **18.** 156.2 lbs. **19.** 3.125 lbs.; 9.375 lbs. **20.** 25 lbs.

Page 73

1. 700 cm./sec. **2.** 3223 grams. **3.** 7.4 lbs.; unstable. **4.** 1007 grams. **5.** 80 ft./sec. **6.** 26.1 ft./sec.; 11.7 revolutions. **7.** 5 ft./sec.^{2}; angle whose tangent is 5/32. **8.** 4.3 inches. **9.** 3.43 lbs. change in weight. **10.** 42.2 lbs.

Page 77

1. 51 grams; 0.0102. **2.** 1 lb.; 1 lb. **3.** 10 lbs.; $v = 2.53$ ft./sec., or 4.97 sec. per revolution. **4.** 4898 grams. **5.** 22.6 ft./sec. **6.** 200 lbs. **7.** 400 lbs. **8.** 823 lbs.; 0.316. **10.** 4.8 lbs.; 7.8 lbs. **11.** 0.167. **12.** 2000 lbs.

Page 90

1. 250 lbs.; 31,250 ft.-lbs. **2.** 81.6 kg.; 0.005 sec. **3.** 90 lbs.; 0.655 H.P. **4.** 10.9 H.P.; 16.35 H.P. **5.** 825 lbs. **6.** 141.8 min. **7.** 626 m./sec. **8.** 800 ft.-lbs. **9.** 5655 ft.-lbs. **10.** 0.582 H.P. **11.** 0.756 H.P. **12.** 20 lbs.-ft.; 1.14 H.P. **13.** 2.02 H.P. **14.** 2400 lbs.; 7680 ft.-lbs. **15.** 0.96 H.P. **16.** 332 lbs. **17.** 6; 72.9%. **18.** 30,000 ft.-lbs.; 20,000 ft.-lbs.; 15; 66.7%. **19.** 876 ft.-lbs.

Page 107

1. 0.000145; 0.00174; 0.1047. **4.** Turns to right. **5.** 98,000,000; 9.88 cm./sec. **6.** Torque 0.167 lb.-ft.; acceleration 0.222 radian/sec.^{2}; 7.52 sec. **7.** 3.55 radians/sec.; 1.42 revolutions. **8.** 118.3 cm./sec. **9.** 7.9 sec. **10.** 25.1 sec. **11.** 725 cm.

Page 116

1. 0.002 lb. **2.** 6.12×10^{-11} (using grams). **3.** 0.000017 gram. **4.** 2.02×10^{19} kg. **5.** If radius of moon is stated as 3.5×10^8 cm., the force is 2.84 kg.; the true radius is 1.7×10^8 cm. and the answer is 12.0 kg. **6.** Yes. 0.0000035 gram. **7.** 6.08×10^{27} grams.

Page 126

1. 93 m., or 305 ft. **2.** 38.8 cm. of mercury; 2.63 cm. **3.** 1.45 mm. **4.** 1.27×10^9 (using grams and cm.). **5.** 2.5% change.

Page 137

1. 100 dynes. **2.** 0.45 of its weight. **3.** 2.55 cm. **4.** Small one gets smaller. **5.** 5960. **6.** 0.298 mm. **9.** Yes. **10.** No.

Page 146

1. 6.5×10^{11} cm.; 15,600 cm.^{2} **2.** 324 m./sec. **3.** 2.7×10^{19}. **4.** 4.47×10^{17}. **5.** 1346. **6.** 2.7×10^{15} sec., or 8.6×10^7 years. **7.** 24 cm. diameter; 183 meters. **8.** 3.55×10^{10}.

Page 161

2. −40°. **3.** −17°.8 and 21°.1 C.; 1112° and 14° F. **5.** 15° C. **6.** 1000.64 ft. **7.** 1.0036 cm.^{3} **8.** 2.49865. **9.** 143° C. **10.** 13.355. **11.** 101.179. **12.** 100.809 cm. **13.** 0.918 cm.^{3} **14.** 33.3 cu. m. **15.** 60/263. **16.** 1.19 grams. **17.** 112 cm.; makes difference of 3 in 1000. **18.** 46.3 lbs./in.^{2} (gauge). **19.** 268 cu. m. **20.** 1.96 kg. **21.** 374 grams.

Page 173

1. 39.6 liters. **2.** 0.103 (If the rise is given as 8° C., the answer is 0.0307). **3.** 19.2; 0.192. **4.** 5°.16 C. **5.** 1,500,000 cal. **6.** 0.088. **7.** 270. **8.** 79°.8 C. **9.** 19,600,000 ergs; 5°.86 C. **10.** Lead bullet, 31.9° C. **11.** 2°.93 C. **12.** 1.12° C.

Page 184

1. No; less error with water. **2.** 0.48. **3.** Copper kettle, 24,000,000. **4.** In the first printing of this book, this problem should be omitted, as it is a duplicate of an earlier one. In later printings, the problem deals with the

flow of heat through the walls of a house; answer 28.8 millions of calories per hour. **5.** 8.57 millions. **6.** 1035°, 1282°, 1927° C. **7.** 1.24 times.

Page 207

2. Dissolved air is removed by heating. **5.** 1027 gr. **6.** 7.75 gr. **7.** 0° C.; 112.5 grams ice left. **8.** 0° C.; 37.5 grams ice left. **9.** 100 grams ice left. **10.** 991 grams. **11.** 0.27. **12.** 50 min. **13.** 348 grams. **14.** 20 grams ice left. **15.** $L = 5.4$.

Page 226

1. 15 times that of coal. **2.** If coal at $15.00 a ton, electricity costs 10 times as much. **3.** 30,460,000; 8.46; 11.3. **4.** 20%. **5.** 0.66 ton. **6.** 54.9 KW. **7.** 25 : 38. **8.** 38.4 : 45.4. **9.** 0.955 H.P. **10.** 823 H.P.

Page 238

1. 500, 750, 1000, etc. **2.** (*a*) 0; (*b*) 20 cm./sec.^{2}; no difference. **3.** 0.633 sec. **4.** 6.9 lbs. change. **5.** 1.11 sec. period. **6.** 2.828 ft./sec. **7.** 238.8 cm./sec. **8.** Gains; 3.68 sec.

Page 263

1. 200 cm.; 1.7 cm. **2.** 137.5; 412.5; 687.5; etc. **3.** 1700, 3400, 5100, etc. **5.** 31.9 ft./sec. **8.** $T°(\text{abs.}) = 288\,(1 \pm n/435)^2$ where n = no. of beats per sec. **9.** If the problem states that the heaps are 50 cm. apart, the answers are 100 cm; 340 per sec.; 120 cm.; 408 m./sec. If it states that the heaps are 20 cm. apart they are 40; 850; 120; 1020. **10.** 60 : 57; increases. **11.** 504; 496. **12.** 4492 meters.

Page 283

1. 384. **2.** One revolution in 3/4 sec., or twice as fast, three times, etc. **3.** 21.8 vibrations per sec.; the listener is in the car. **6.** Time changes from 4.17 to 1.25 sec. **7.** 50%. **8.** If problem says 1.6 sec. answer is 75 ft.^{2}; if 1.5 sec., answer 450 ft.^{2}

Page 301

4. Touch the end of one to the middle of other. **5.** Intensity = 0.46 unit. **6.** 7.07 dynes. **7.** 3.06 dynes. **8.** 0.1 unit, southward. **9.** 0.1452. **11.** 160 dyne-cms. Torque = MH. **12.** $m = 103$.

Page 325

1. 0.04. **2.** 2.25 dynes. **3.** 0.31 dyne. **4.** 15 dynes; 7.5 dynes; 1 dyne. **5.** −19.6. **6.** 12.14 dynes, along diagonal. **7.** 0.19 dyne. **8.** $Q = 333.3$; $V_1 = 33.3$; $V_2 = 66.67$. **9.** $Q_1 = 50$; $Q_2 = 200$; $Q = 250$; $C = 25$. **10.** $Q = 200$; $V = 200$. **11.** $Q = 24{,}000$; $Q_1 = 4000$; $Q_2 = 8000$; $Q_3 = 12{,}000$. **12.** V : 180, 90, 30; Q : 9000. **16.** 6370. **17.** 5,096,000 ergs.

Page 349

1. None. **2.** 0.0000393. **3.** 1.005 ohms. **4.** 0.74 mm. **5.** 0.386 ohm. **6.** 0.923; 1.2; 1.333; 1.555; 1.714; 2; 2.222; 3; 3.714; 4; 4.333; 5; 5.2; 6; 7; 9. **7.** 8 ohms; 4 volts. **8.** 20 volts. **9.** Shunt with 0.51 ohm. **10.** 24 and 2 amps. **11.** 1.0101; no; too low resistance. **12.** (*a*) Shunt with 0.202 ohm; (*b*) 980 ohms in series. **13.** 0.235 volt; 0.0784 amp. **14.** 7 ohms; 7 volts. **15.** 0.727 volt.

Page 369

1. 0.01 ohm. **2.** (*a*) 2.0 volts; (*b*) 2.2 volts. **3.** 1.5 volts. **4.** 12 amperes; 1.2 volts. **5.** 13.5 volts; 3.0 volts; 1 ampere. **6.** 9.56 ohms in series. **7.** 21.91 ohms. **8.** (*a*) 8 volts; (*b*) 2 volts (neg.).

Page 376

1. 6.33 min. **2.** 220 ohms; 55 watts; 13.16 cal. **3.** If the resistance is stated as 0.5 ohm: 100 volts; 210 volts; 20 KW; line resistance is too high. If resistance is 0.1 ohm, answers are 20 volts; 130 volts; 4 KW. **4.** 95.7. **5.** 480 volts; 1 KW. **6.** In the first printing of this book this problem should be discarded, as it is a duplicate; in the later printings this problem refers to the water evaporated in a storage battery; answer 7.18 grams. **7.** 5 ohms; 119.6 cal. **8.** 0.25 amp.; 1 amp.; one lamp gives (*a*) 3.29 cal.; (*b*) 13.15 cal. **11.** 22.5 watts; heat.

Page 384

2. From 58,800 to 198,000. **3.** 2290 to 24,600.

Page 398

1. 0.15 volt. **2.** 1.5 volts. **3.** 24,000 lines; 0.0048 volt. **4.** 16,000 volts. **5.** 0.2 volt.

Page 410

3. 108 volts; 55 amperes; 1447 calories. **4.** 12 volts. (*a*) direct; (*b*) direct; (*c*) according to B–H curve. **5.** 105 volts; 21 ohms. **6.** 750 R.P.M. **7.** 66.7 volts. **8.** 58.1 amp.; 108.7 volts. **9.** 88%. **10.** 55 amperes. **11.** 100 volts; 60 amperes. **12.** 106.24 volts; 245 amp.; 468 amp. **13.** 5 KW; 406,000 gr.-cm. **14.** 3 amp.; 4.3 cal./sec.; 312 watts; 16.33 ohms.

Page 431

2. 5000 volts; 0.2 amp.; 1000 watts. **3.** 40 volts; 1 amp.; 2 KW. **4.** 10,000,000 lines, counting each conductor separately; 0.02 sec.; 5 volts; 80 volts. (This is 0.636 of the maximum value; the R.M.S. value is 0.707 of the maximum.) **5.** 2.96 amp. **6.** 0.866. **7.** 4.79 KW.

Page 438

1. 101.3 henries. **2.** 1.875×10^8. **3.** 7500 meters.

Page 450

1. $v = 2.66 \times 10^8$ cm./sec.; energy $= 3.18 \times 10^{-}$ ergs. **2.** Energy $= 1.59 \times 10^{-7}$ ergs; $v = 1.9 \times 10^{10}$. Actually, m would increase, and v be less.

Page 473

4. 1.66 ft.-candles. **5.** 128 c.p.; lowered. **6.** 55.5 c.p.; 2.22 ft.-candles. **7.** 82.84 cm. from end.

Page 480

1. 7.4. **2.** 6×10^{14}. **3.** Reduces to one millionth in 338 reflections.

Page 488

1. $n = 2$. **2.** 1.25×10^{10} cm./sec. **4.** Half-length.

Page 503

1. $q = -2.222$ cm. **2.** $q = -90$ cm. **3.** $n = 1.6$. **4.** 9.75 ft. **5.** $R = 12$ cm. **6.** 1.545 **7.** 30 in. **9.** 4.69 cm. **10.** 2 in.; 0.5 in.

Page 521

1. (*a*) -38 cm.; (*b*) $+42$ cm. **2.** Flat cornea; no adjustment required for distant objects. **3.** $q = -8.1$ cm. (from negative lens). **4.** $+30$ in. from neg. lens. **5.** $q = +1.11$ cm. beyond second lens. **6.** $q = 9.86$ cm. beyond neg. lens. **7.** $q = +16.37$ cm. beyond neg. lens. **8.** $+1.62$ cm. beyond eyepiece. **9.** 9.3 m. **10.** 27.19 cm. **11.** 25 cm.; 83.3. **12.** 68.67 cm. **13.** $q = 0.449$ in. beyond eyepiece; not satisfactory at all. **14.** $q = 32.3$ in. beyond eyepiece. **15.** $q = 12$. No. **16.** 49 : 64.

Page 567

1. Purples, pinks, etc. are hardest. **2.** Black. **6.** This grating would form ultra-violet spectra, but no visible. **7.** Wave-length 7143 Angstrom units. **8.** Wave-length 4020 is the only region cut out; color straw yellow. **9.** 0.000157. **10.** 21,000; 10,500; 7000; 5250; 4200; etc.